

nature

RING AROUND THE MOON

A new generation of
solar eclipse research

**PRESERVING THE
DIGITAL AGE**
Personal informatics

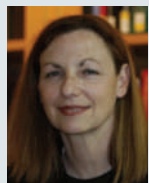
**EVOLUTION OF
IMMUNITY**

The Bredt/T cell
split came early

**SOLAR SYSTEM
EVOLUTION**
A binary role?

EXCLUSIVE VIEW
Looking out to
the universe

Abstractions



LAST AUTHOR

In 1999, researchers reported that increasing production of the Sir2 protein in the yeast *Saccharomyces cerevisiae* extends the organism's lifespan. This and other

findings spurred a flurry of activity and the identification of similar longevity proteins, dubbed sirtuins, in other organisms. But how sirtuins promote long life remains unclear. Shelley Berger of the Wistar Institute in Philadelphia, Pennsylvania, and her colleagues have now uncovered a clue (see page 802). As yeast cells age, Sir2, an enzyme known as a deacetylase, becomes less abundant and so less effective at removing acetyl groups from histone H4 — one of the proteins that package DNA into tight coils — causing the packaging to fall apart. Histone H4 was known to be one substrate of Sir2, but until now, not ageing-specific. Berger talks to *Nature* about her foray into ageing research.

Why did you start this project?

My research focuses on epigenetics, the study of modifications that occur in chromatin — the combination of DNA, RNA and proteins making up chromosomes — and that are transmitted from one generation to the next, affecting how genes function. I thought that yeast might be a good model system for addressing whether changes in chromatin are involved in ageing. Yeast is currently the only organism in which we can easily manipulate histones using genetic techniques.

What did you find?

We were surprised to find that Sir2 protein levels decrease as yeast cells age. Other groups had looked at the levels of Sir2 RNA, and those remain constant. As the protein's levels drop, there is a corresponding increase in acetylation of the amino acid lysine at position 16 of the histone H4 in the chromatin located near the ends of yeast chromosomes.

Does this have implications for higher organisms or therapeutics?

It's likely. Several sirtuins function as histone deacetylases in other organisms, including mammals. And, in mammalian cells, the ends of chromosomes, or 'telomeres', are known to become unstable and shorten with age. In terms of therapeutics, we identified an acetylase, Sas2, that has the opposite function to Sir2. The hunt has been on for activators of sirtuins as longevity drugs. But with Sas2, perhaps we can start thinking about inhibitors.

Had you worked on ageing before?

No. Coincidentally, I was a postdoc in Leonard Guarente's lab at the Massachusetts Institute of Technology, where the Sir2 pathway was first found to increase lifespan. But I left just as Lenny was starting studies on ageing. ■

MAKING THE PAPER

Phil Baran

Natural clues offer fresh route to chemical synthesis.

Phil Baran still remembers the date that he ventured into new territory. On 12 January 2007, an editor at *Nature Chemical Biology* contacted him to ask whether he might write a review article on terpenes — a large class of naturally occurring molecules that have applications in everything from perfumes to pesticides. Until then, Baran, an organic chemist at the Scripps Research Institute in La Jolla, California, had worked only on nitrogen-containing compounds called alkaloids. But he agreed to take a look at terpene synthesis. Two-and-a-half years later, having learned to mimic a process that nature has taken eons to perfect, he is setting precedents in the terpene field.

Natural terpenes are built in a two-stage process: first, enzymes form the molecule's hydrocarbon backbone, which Baran likens to an undecorated Christmas tree; second, oxygen-containing groups are added to the backbone, like ornaments to the tree's branches, to create a wide array of molecules. Although living organisms can churn out massive quantities of these chemicals, chemists have been frustrated in their efforts to synthesize them in amounts sufficient for commercial use.

Having pored over hundreds of manuscripts spanning 50 years' worth of terpene production, Baran realized that most approaches involved a large number of steps and a complex building process. This, he thought, would make it difficult to make large quantities of any compound. "At a molecular level we're pretty good at making Christmas trees, but not so good at putting the ornaments on," he says.

He reasoned that the key to success might be to find a way to simplify terpene synthesis by mimicking natural mechanisms. To do so, he and research associate Ke Chen applied a twist to 'retrosynthetic analysis' — a commonly used



approach for working out how to build a molecule. The process normally starts with a single target molecule and works backwards, breaking apart strategic bonds to make simpler and simpler molecules, to help outline possible synthesis strategies.

But terpene synthesis is complicated by the fact that highly oxygenated molecules have functional groups that can easily react with one another or prevent key reactions from proceeding. So chemists often have to remove a recently added functional group before they can add a different one, and restore it later — an approach that is obviously wasteful.

As described on page 824, Baran and Chen set out to develop a more systematic process using the 'eudesmane' terpene class as a proof-of-principle. In the reverse direction, or planning stage, each disconnection leads to a set of molecules called redox isomers (rather than just one molecule, as in conventional retrosynthetic analysis). In each set, the molecules have the same number of ornaments, although in different locations, and all have one fewer ornament than in the previous set. This process continues until a set of bare carbon skeletons remain.

During synthesis — the 'forwards' direction — the researchers started by making large quantities of a terpene with a single hydroxyl group, which Baran likens to the angel at the top of the tree. They then added ornaments in a methodical and controlled way, without having to remove one thing before putting another on. In this way, they synthesized five types of eudesmane in far fewer steps than had been done using conventional synthetic approaches.

The molecules synthesized by Baran and Chen are simpler than commercial terpenes such as the anticancer drug Taxol. But Baran says that their methodology lays the foundations for synthesis of more complex terpenes. ■

FROM THE BLOGOSPHERE

The blogosphere was a-twitter last week with news of the first scientific experiment to be conducted on Twitter, the Internet-based instant-messaging service (<http://twitter.com>). The experiment may lack rigour, but it does encourage public participation in the scientific process.

On The Great Beyond blog, assistant news editor Richard Van Noorden reports that

psychologist Richard Wiseman will use Twitter to test 'remote viewing' — the psychic ability to identify a distant locale without being shown or told what it is (<http://tinyurl.com/kpjf76>). Wiseman will 'tweet' from an undisclosed location, and then allow those following him on Twitter to vote on five possible locations posted as photos on a website.

If the voting were

independent, Wiseman, at the University of Hertfordshire, UK, calculates that getting a majority of voters to 'hit' on his location by chance in at least 3 out of 4 trials would be 1 in 37. However, the nature of social networking means that voting will not be truly independent. The results from four such trials are expected to be published on Wiseman's blog this week (<http://tinyurl.com/kls85q>). ■

Visit *Nautilus* for regular news relevant to *Nature* authors ▶ <http://blogs.nature.com/nautilus> and see Peer-to-Peer for news for peer reviewers and about peer review ▶ <http://blogs.nature.com/peer-to-peer>.

Unjust burdens of proof

English libel law adversely affects every publisher and website host whose content can be read in the United Kingdom. It must be changed.

In 2004, the book *House of Bush, House of Saud* reached the best-seller list in *The New York Times*. It was on sale in bookshops in France, Germany and Japan. But it was not published in Britain. It thus joined the ranks of books — *Nobel Dreams* by Gary Taubes was another — that, one can only infer, were in effect censored owing to the reluctance of publishers to risk a libel suit under English law.

The reason? At the heart of the issue sits the burden of proof of unjustified defamation. In the United States, aggrieved ‘claimants’ who believe that a book, magazine or newspaper has significantly damaged their reputations without justification generally have to prove to the courts that the allegations are false. In Britain, the author and publisher (assuming both are sued) have to prove that the allegations are true. Also critical to the outcome is an interpretation by the court of the meaning in the mind of a reasonable person of what was originally published.

From time to time, *Nature's* journalism has stimulated people to get their lawyers to send us threatening letters. Because we have been able to show that our reporters tackle issues of genuine public interest, do their homework and give people the opportunity to reply to allegations within our stories, such threats have generally been rebuffed.

Simon Singh, a leading independent science writer, was less fortunate. In an article in *The Guardian*, he tackled an issue of genuine public interest: whether or not there is any evidence that chiropractic treatment is effective, as some chiropractors have suggested, for various childhood conditions, including asthma. The British Chiropractic Association (BCA) claimed that he had defamed its reputation and threatened to sue him for libel. According to Singh, the BCA did not take up an offer by *The Guardian* of space in which to respond.

Again according to Singh, he was expressing his own opinions and was not accusing the BCA of being dishonest. But last week a judge controversially decided that Singh's language amounted to factual statements (rather than opinion) and that the meaning of Singh's

article was that the BCA was deliberately using unfounded claims. This meaning makes defending a libel case extremely difficult, as he would have to prove that that was indeed the case, even though it was not the meaning that he intended. So he has decided to appeal against the ruling on meaning.

The campaign that was launched in support of Singh makes the valid point that debates about scientific claims should be conducted in the open rather than pursued in the courts. In a statement, the BCA claims to support this ethos and that it was only opposing Singh's characterization of its integrity. However, many observers, including *Nature*, believe that its resort to the courts is tantamount to suppression rather than toleration of debate.

There is no reason to give a special privilege to science within libel law. What is required is that English law be made not only consistent with that of other countries, but also with Article 10 of the European Convention on Human Rights, which protects the right of freedom of expression. The letter of English law may not breach the latter, but the threat of the costs involved, which can be huge, and the burden of proof are enough to scare many writers and publishers into a restraint of freedom of expression.

Anyone who believes that this is an issue only for those living in Britain could not be more mistaken. It affects any online content that can be read in the United Kingdom. Wisely, there is legislation in the United States that, in some instances, protects its citizens from English court judgements.

The English law needs to be changed. There urgently needs to be a review of the burden of proof and a restriction of the costs that can reasonably be awarded to whoever wins the case. Only then can Britain claim to be a country that supports free speech. ■

“There is no reason to give a special privilege to science within libel law.”

Watch your back

The H1N1 flu epidemic is not the world's only disease threat.

If you are a health official facing two highly contagious diseases — one that is already killing dozens of people, and another that threatens to kill people by the thousands or millions, but hasn't yet done so — how do you allocate your resources? The answer can be a difficult balancing act, as the situation in China is showing.

At the moment, China's top health priority is the potential threat of pandemic H1N1 swine flu virus. Over the past weeks China has had 108 confirmed cases of H1N1, so far without any deaths. But officials

fear that there could be enormous numbers of fatalities in the coming months if the virus spreads through the country's 1.3 billion people. Responding to World Health Organization (WHO) warnings and global concerns, China has acted decisively and sometimes excessively, possibly reflecting the country's delayed response to severe acute respiratory syndrome in 2003. Medical teams have been sent on to planes to measure the temperature of every passenger aboard, despite evidence saying that such measures are not particularly effective, and Mexicans have been quarantined seemingly just because they come from the country where the H1N1 outbreak was first detected.

At the same time, however, children around the country are falling ill because of hand, foot and mouth disease (HFMD) — an illness that has become an increasingly serious problem in many parts of Asia since a 1997 outbreak in Malaysia killed 29 people. Especially worrying is

that, unlike diseases such as dengue fever or malaria, but like H1N1, HFMD can be transmitted directly from human to human; no animal reservoir is required. Cases start out with flu-like symptoms that can develop into characteristic sores in the mouth and rashes on the palms and soles of the feet. Infections can also lead to meningitis and swelling of the brain, possibly resulting in long-term neurological damage (L.-Y. Chang *et al.* *N. Engl. J. Med.* **356**, 1226–1234; 2007) or even death. The vast majority of reported cases and almost all of the serious cases are in patients under 5 years old.

In 2008, more than 500,000 HFMD infections were reported in China, and 200 people died. This year, as the disease moves through its peak season, similar numbers look possible. There is currently no vaccine to prevent it and no drug to cure it. The best safeguard is good hygiene: HFMD transmission is mainly via saliva and faeces. But even Singapore, renowned for its cleanliness, has had 8,896 cases this year, and had had even more by this time last year. There have also been cases this year in Taiwan, Hong Kong and South Korea.

Most of the serious HFMD cases seem to be caused by enterovirus 71 (EV71), first isolated in California in 1969. EV71, in turn, is related to poliovirus — a fact that has some experts concerned. EV71's recent pattern, with a few isolated outbreaks that are building in frequency,

is similar to that seen before the twentieth-century polio epidemics that killed or crippled tens of millions of people around the world (J. F. Modlin *N. Engl. J. Med.* **356**, 1204–1205; 2007). Although it is far from certain that EV71 will follow a similar course, the virus does seem to have become increasingly virulent in China since 2004, possibly because of genetic changes (see *Nature* **458**, 554–555; 2009).

Thus the balancing act: China, the WHO and all of the other health organizations are to be commended for mounting a vigorous response to H1N1 (see page 756). But given the ongoing damage being caused by EV71, and its potential for large-scale, person-to-person spread, they should also be mounting a substantial effort to understand and counter that virus. And indeed, the WHO had a meeting on EV71 in Beijing earlier this year.

Yet that good start is not enough. WHO representatives play down the importance of EV71, and there is little information about the virus on the WHO website. One EV71 specialist laments: "All resources seem to be going to influenza." And in the meantime, many doctors in China still misdiagnose HFMD, partly because the virus and its symptoms change, and partly because there are still no good tests to identify it. More leadership, more education, more awareness and much more research on drugs and a vaccine are essential. ■

The sharing principle

Journals and funders must insist that genetically modified mouse strains are fully available.

Back in 1996, human-genome scientists signed up to the Bermuda agreement to share their data without delay. Since then, the sharing principle has entered the mainstream — it now applies to all genomic data generated using public funding, as well as to all the relevant resources cited in publications.

But this principle is not universally observed for genetically modified mice, designed as vital resources in the quest to unpick basic biological mechanisms or to model human disease. The size of the problem is unclear, but existing surveys, combined with extensive anecdotal evidence, suggest it is substantial. In April 2006, for example, scientists at the US National Institutes of Health found that nearly 4,000 unique mice strains had been created, yet barely 700 had been placed in a repository.

Some scientists say they do not have the time nor money to breed and distribute their mice, or even to send the animals to publicly funded mouse repositories such as the European Mouse Mutant Archive in Europe, the Jackson Laboratory in Bar Harbor, Maine, and RIKEN BioResource Centre in Ibaraki, Japan, which would do those chores for them. Others claim that the careers of young and vulnerable researchers (or old and vulnerable researchers) could be harmed if they lost their exclusive access to a resource they made for their own research projects. Or, they say their institution's technology-transfer offices or companies will not let them part with mouse strains that could perhaps be made to turn a profit.

Such attitudes were noted with concern last month at a workshop in Rome hosted by CASIMIR, a European Union project to

coordinate and sustain mouse resources internationally (see background documents at <http://tinyurl.com/l2hyvq>). The workshop brought together representatives from funding agencies, publishers and the mouse repositories from Europe, the United States and Australasia. They concluded that the sharing problem urgently needs resolution — not least because international projects to systematically generate mouse lines deficient in each gene in the genome will generate thousands of new strains in the next five years or so.

To solve the problem, however, journals and funding agencies must take a tougher line. The *Nature* journals are among the very few actually requiring that authors use established public repositories wherever possible as a condition of publication. Most journals simply 'encourage' their authors to make mice used in their publications freely available to other laboratories, or 'suggest' that the mice be deposited in repositories. Funding agencies similarly prefer such cajoling terms as 'encourage' in their policies on sharing mouse resources, and rarely police the outcome.

Journals should now require researchers to place their mice in repositories as a condition of publication. And funding agencies should require repository plans to be included in all grant applications that are likely to generate new mouse strains. Part of the grant money should be reserved for this task and final reports or evaluations of the grants should refer to the repository used. The repositories themselves should help the journals and funding agencies by finding a way to generate a unique accession number for each mouse strain.

The sharing principle allows biology to progress efficiently. It avoids duplication of effort and allows different laboratories to use the same tools. It is essential that scientists sign up to it. Sharing mice has never been easier — the repositories around the world are efficient and professional, and they are coordinated. Just a few changes in the modus operandi of key institutions could ensure that the makers of mice will have no possible excuse not to use them. ■

RESEARCH HIGHLIGHTS

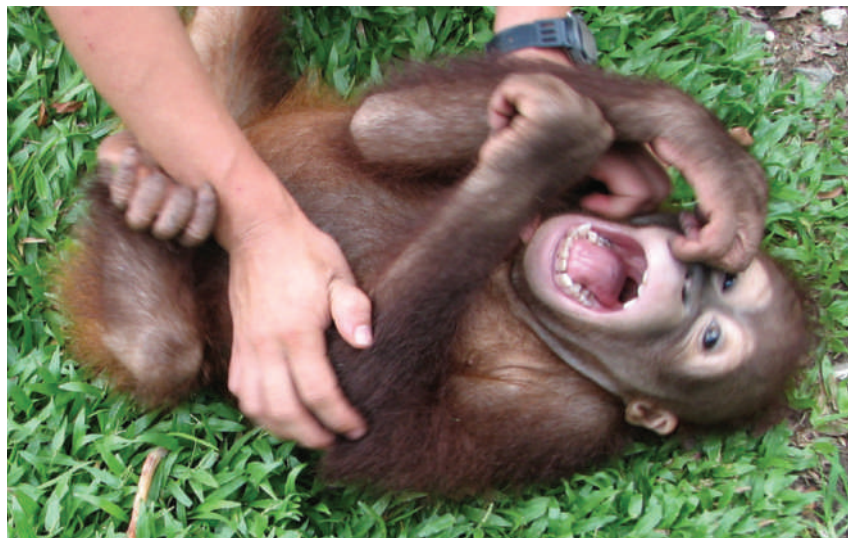
Tickle tree

Curr. Biol. doi:10.1016/j.cub.2009.05.028 (2009)

A study of recordings made while researchers tickled infant humans and juveniles of several species of ape shows that the relationship between different species' laughter closely matches an independent family tree derived from genetic differences.

Biologists have debated whether human laughter, which sounds distinct from that of apes, consists of true innovations, or simply continues evolutionary pathways pioneered by ape-like ancestors. Marina Davila Ross of the University of Portsmouth, UK, and her colleagues conclude from their analysis of the recordings that tickle-induced laughter has a common evolutionary origin in apes and humans. They add that the analysis could have implications for the evolution of human speech.

For a longer story on this research,
see <http://tinyurl.com/phh2m9>



M. WESSELS

ASTRONOMY

What's handed down

Astrophys. J. **698**, L37–L41 (2009)

In the unusual chemistry of a bright star, researchers have detected the echo of the Universe's first stars, massive but short-lived giants that existed about 13 billion years ago.

BD+44°493 is an extreme example of a star enriched in carbon but depleted in iron, say Hiroko Ito of SOKENDAI, the Graduate University of Advanced Studies in Tokyo, and her colleagues. The high amount of carbon could not have been produced by the star's own fusion synthesis, say the researchers, nor is it likely to have come from a companion star or a previous generation of rapidly spinning stars. They thus surmise that it originated in the turbulent, black-hole-producing supernovae of the Universe's first stars.

ECOLOGY

Evidence of emperors

Glob. Ecol. Biogeogr. doi:10.1111/j.1466-8238.2009.00467.x (2009)

Emperor penguin colonies are hard to locate because they are distributed over huge swathes of sea ice, much of which is present only in the most inhospitable months of the year.

Peter Fretwell and Philip Trathan of the British Antarctic Survey in Cambridge, UK, have found a way to spot breeding colonies without clambering around in the cold; they look at satellite images for the large faecal stains (pictured right) created by the emperors (*Aptenodytes forsteri*).

Their survey has located 38 colonies, including 10 new sites and 6 that were

recorded as being in the wrong place. Six other previously recorded colonies were missing.

PLANETARY SCIENCE

Pressure drop

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0809436106 (2009)

The Sun is growing brighter with age, and some argue that this will limit the lifespan of Earth's biosphere. The hotter the Sun, the more readily carbon dioxide will be removed from the atmosphere by chemical weathering; once levels fall too low, photosynthesis will cease and the biosphere will be radically diminished.

King-Fai Li and his colleagues at the California Institute of Technology in Pasadena offer a way out of this trap with the help of a simple model. A drop in overall atmospheric pressure, achieved perhaps through biological sequestration of nitrogen, would weaken the atmospheric greenhouse effect. This would

make the world cooler for a given level of CO₂. As a result, the weathering effects would be less marked and CO₂ levels could stay higher for longer. This might extend the biosphere's remaining lifespan from around 1 billion years to something like 2.3 billion years.

The authors conclude that atmospheric pressure is also likely to co-regulate the climate of planets elsewhere in the Universe. If this means that the lifetimes of biospheres are longer in general, it might increase the chances of intelligent life evolving, and the chances that that life coincides with life on Earth.

OCEANOGRAPHY

Glowing oceans

Biogeosciences **6**, 779–794 (2009)

Building on earlier research into global assessments of ocean productivity, a team led by Michael Behrenfeld of Oregon State University in Corvallis reports the first satellite-based evidence linking phytoplankton fluorescence to iron stress in surface waters.

Using images collected by the MODIS sensor aboard NASA's Aqua satellite between 2003 and 2007, the team found that high fluorescence levels were highly correlated with low iron levels in surface waters.

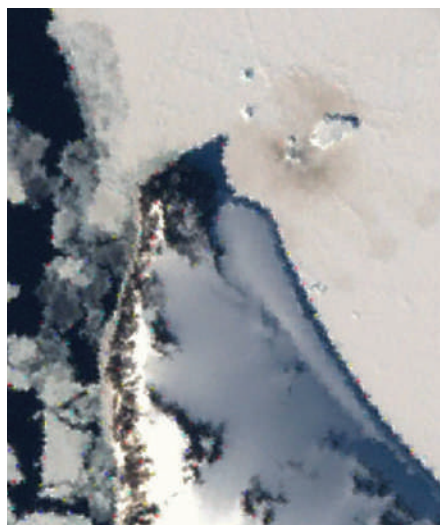
Researchers say the technique could be used to monitor iron stress in phytoplankton and to analyse iron-fertilization experiments.

NEUROGENETICS

Huntington's toxic trigger

Science **324**, 1327–1330 (2009)

Uncontrolled movements are symptomatic of Huntington's disease, which is caused by mutation of the huntingtin gene. Mutant



LIMA/USGS

huntingtin protein (mHtt) is expressed throughout the bodies of carriers, but mHtt-associated cell death occurs selectively in certain brain regions, including the striatum, which is important for movement.

A team led by Solomon Snyder at Johns Hopkins University School of Medicine in Baltimore, Maryland, found that Rhes, a protein located mainly in the striatum, binds to mHtt and attaches a small protein to it. This modification increased cell death in cultured cells, suggesting that Rhes interacts with mHtt to cause localized neurodegeneration.

ECOLOGY

Biomes bounce back

PLoS ONE 4, e5653 (2009)

Ecosystems damaged by human activities such as agriculture and oil spills may be quicker to recover than was thought. Rather than taking centuries, or even millennia, to convalesce, Holly Jones and Oswald Schmitz at Yale University found that even severely damaged ecosystems could recover within decades.

The duo reviewed 240 independent scientific studies conducted between 1910 and 2008 that examined the recovery of large ecosystems from damaging human activities and natural disturbances such as hurricanes. They discovered that the longest average recovery time — found to be taken by forest ecosystems — was no more than 56 years. Deep-ocean ecosystems took only about 5 years.

HIGH-ENERGY PHYSICS

Muonium gets real

Phys. Rev. Lett. 102, 213401 (2009)

Muons are heavy, unstable cousins of electrons that are produced by high-energy collisions of more commonplace particles.

Since the 1960s, high-energy physicists have observed atomic-like 'muonium' particles made up of an antimuon and an electron, or 'muonic atoms' in which a muon replaces an ordinary atomic electron. But nobody has seen 'true muonium', made up of a muon and an antimuon. Now Stanley Brodsky of the SLAC National Accelerator Laboratory at Stanford University in California and Richard Lebed of Arizona State University in Tempe suggest that physicists could spot the stuff if they just look in the right place.

They show that true muonium can briefly appear in colliding beams of electrons and their antiparticles, positrons. In fact, the authors say, true muonium may have already been made in modern electron-positron colliders.

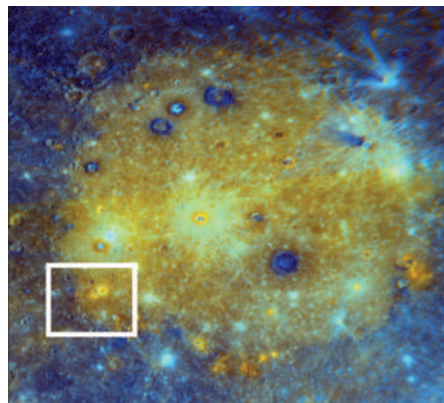
PLANETARY SCIENCE

Mercurial Mercury

Earth Planet. Sci. Lett. doi:10.1016/j.epsl.2009.04.037 (2009)

Baked by the Sun and blasted by impacts, Mercury is thought to have lost much of its volatile content — such as water vapour and carbon dioxide — early in its history. But its interior may have been bubblier than thought, according to Laura Kerber of Brown University in Providence, Rhode Island, and her colleagues. Using observations from the MESSENGER spacecraft, the researchers spotted the remnants of a volcano that they say was driven by explosive magma with high volatile concentrations.

The large size of the volcanic dome (marked by box below) suggests a volatile content comparable to that seen in eruptions on Earth. The discovery favours the idea that Mercury formed from volatile-rich planetesimals from the outer Solar System that migrated inwards.



CANCER BIOLOGY

Tumour clue

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0900351106 (2009).

The blood-pressure drug losartan could fight some breast cancers, studies in mice suggest.

Arul Chinnaiyan of the University of Michigan in Ann Arbor and his colleagues analysed 3,200 experiments that characterized gene-expression patterns in breast cancer tumours. The search revealed that a gene called *AGTR1* was expressed at higher than normal levels in 10–20% of the cancers.

Human mammary cells that express *AGTR1* at abnormally high levels became more invasive when stimulated with angiotensin II, which activates the *AGTR1* protein. Meanwhile, losartan, a treatment for high blood pressure that inhibits *AGTR1*, reduced this response. It also reduced tumour growth by 30% in mice implanted with *AGTR1*-overexpressing breast cancer cells.

JOURNAL CLUB

Gail Christeson

University of Texas, Austin, USA

A geophysicist ponders the mysteries of intraplate earthquakes.

During my first semester at college, I attended a lecture describing plate tectonics, and immediately knew that geophysics would be my major and hopefully my career. Subsequent lectures, textbooks, journal articles and, later, my own research educated me about how elegantly plate tectonics explains the processes that control the locations of most earthquakes.

However, some of the largest-known North American earthquakes — including those of 1811–12 famed for ringing church bells in Boston and changing the course of the Mississippi River — are associated with the New Madrid Seismic Zone (NMSZ) in the Southern and Midwestern United States, far from known plate boundaries. So what causes these events? Eric Calais of Purdue University in Indiana and Seth Stein at Northwestern University in Illinois present some surprising results from an examination of Global Positioning System (GPS) data from the region (E. Calais and S. Stein *Science* 323, 1442; 2009).

Previous studies found that the NMSZ was moving at a different rate and in a different direction from the North American Plate, implying that strain would steadily accumulate until released by a large-magnitude earthquake. But, incorporating three years' worth of extra GPS data, Calais and Stein found motions indistinguishable from those of the North American Plate, corresponding to extremely low strain rates. It is not clear what the underlying processes causing the NMSZ earthquakes are. Is strain accumulation variable over time in intraplate settings? What are the implications for hazard prediction?

The results leave me perplexed, but oddly comforted — there are plenty of mysteries left for the next generation of geophysicists. And perhaps one day a theory will elegantly explain intraplate seismicity, just as plate tectonics did for interplate seismicity.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS

Avian influenza aided readiness for swine flu

Despite gains from threat of bird flu, pandemic preparedness is patchy.

With the World Health Organization discussing whether to declare, as *Nature* went to press, that a swine influenza pandemic has arrived, doctors, scientists and government officials say the enormous preparedness efforts of recent years have aided the world's response to the virus.

"Clearly, the global preparedness for dealing with an influenza outbreak is much better now than it was five or six years ago," says virologist Malik Peiris of the University of Hong Kong.

But there have been some hiccups, due largely to the mismatch between the pandemic scenarios envisaged and the one that has arrived. Most nations had prepared for a pandemic sparked by the deadly A(H5N1) avian influenza virus, which kills up to 60% of those infected, but the A(H1N1) swine flu virus epidemic has been much less severe so far.

"Many nations built their plans around the idea that a pandemic strain would evolve in

southeast Asia, that we would recognize it early, and that we would be able to contain it," says Eric Toner, a physician and preparedness analyst with the University of Pittsburgh Medical Center's Center for Biosecurity in Pennsylvania. "None of that turned out to be true."

Global pandemic planning efforts were ramped up after severe acute respiratory syndrome (SARS) swept through Asia in 2003, and the deadly H5N1 strain emerged in the same year. Local, national and international entities involved in flu response are now better coordinated, observers say.

That was evident with regard to the current situation, as officials — for instance, at the US Centers for Disease Control and Prevention in Atlanta, Georgia — delivered open and consistent public messages about the seriousness and uncertainty of swine flu infections, says Georges Benjamin, executive

"The idea that one could trap it at the early stages was certainly not successful."



director of the American Public Health Association in Washington DC. That's in contrast to 2001, when some conflicting messages were delivered during the anthrax attacks.

In some countries, pandemic planning has led to stockpiled antiviral medication (see 'Preparedness by the numbers') being distributed to treat people with swine flu and their contacts. And the boost in research spending on epidemiological modelling and laboratory studies has enabled scientists to understand the H1N1 virus in record time, prepare to manufacture more vaccine faster, make more informed decisions on how to intervene in the virus's spread, and understand and share information on its genetic characteristics.

"We got an extraordinary laboratory understanding of this virus in a very short period of time, and that is the scientific 'wow' of what really happened in this outbreak," says Benjamin.

But the divergence between the anticipated pandemic and the one that has actually arrived has caused some glitches. For instance, the swine flu virus was not detected in time to contain it, as had been envisaged in some scenarios involving avian flu. "By the time the surveillance systems picked [the virus] up, it was already spreading way out of control, so the idea that one could trap it at the early stages was certainly not successful," says Peiris.

As swine flu began to spread throughout North America, which had hoped to have weeks of warning before a virus arrived from Asia, officials were initially confused about how to implement response plans. For instance, swine flu symptoms were milder and less severe than anticipated — impossible to distinguish from seasonal flu — and hospitals and health departments were swamped with people seeking tests for swine flu. A report issued on 4 June

PREPAREDNESS BY THE NUMBERS

Number of European countries surveyed in 2005 that had a pandemic flu plan: **21 of 29**

Average completeness of the plans*: **54%**

Number of Latin American countries surveyed in 2007 that had a flu plan: **15 of 17**

Average completeness: **48%**

Number of Asian countries surveyed in 2005 that had a flu plan: **6 of 9**

Average completeness: **Not available**

Number of African countries surveyed in 2007 that had a flu plan: **35 of 53**

Average completeness: **36%**

Doses of A(H1N1) vaccines ordered by:

United States: **40 million**

United Kingdom: **90 million**

France: **50 million**

Belgium: **12.6 million**

Finland: **5.3 million**

Australia: **10 million**

Per cent of the population covered by government antiviral stockpiles in:

United States: **25%**

United Kingdom: **50%**

Hong Kong: **285%**

Thailand: **5%**

All nations outside the nine currently capable of producing a pandemic vaccine: **1%**

*Based on the World Health Organization checklist for influenza epidemic preparedness.



Britain has sufficient stockpiles of antiviral drugs to protect 50% of its population.

K. COLLINS/REUTERS

by the non-profit group Trust for America's Health, based in Washington DC, also noted that budget cuts at health departments at all levels slowed the US response, for instance by delaying testing.

"We have to consider what happened in the month of May as maybe a warning of what's coming, and we're working very hard this summer to have countermeasures available if the need for those arises" in the autumn, says Robin Robinson, principal deputy director of the Biomedical Advanced Research and Development Authority within the US Health and Human Services department.

By contrast, Asia had expected an avian pandemic virus to arise in its backyard and was not fully prepared for swine flu from overseas. Most of the flu plans for the members of the Association of Southeast Asian Nations "focus on how to cull chickens", says Hitoshi Oshitani, a virologist with the Tohoku University Graduate School of Medicine in Sendai, Japan. "They may have some

rapid containment plan, but containment is not possible for this current virus," he says.

Some Asian countries have applied containment measures that had little chance of working. China, for example, quarantined foreign travellers with fever symptoms and symptomless travellers in contact with those, a move not backed by scientific evidence, Toner says.

When the virus spread to Europe, some nations found they were not as ready as their thorough plans had led them to believe they were. For instance, Britain's Health Protection Agency, which was responsible for investigating the initial swine flu cases, did not at first have enough staff available to carry out the investigations, says Sandra Mounier-Jack, a health-policy analyst at the London School of Hygiene and Tropical Medicine. And a planned telephone line that was to have been used to help distribute antivirals had not been set up when swine flu appeared.

These are common problems, says Mounier-Jack. "The major gaps are about making the plans operational — making sure that people at local levels know their roles and responsibilities, even in countries which have very good plans and are rich."

And some long-standing challenges for Europe have still not been addressed, such as the piecemeal nature of the response. "We have 27 pandemic preparedness plans in different states of maturity, and now individual ministries of health are all lining up to buy up vaccines," says Albert Osterhaus of the University of Rotterdam in the Netherlands.

But perhaps the biggest global challenge is arriving now, as the Southern Hemisphere enters its flu season and swine flu threatens many of the least-ready countries. "In general, the developing countries are not prepared," Oshitani says. The World Bank has released billions of dollars for preparedness in these regions, but many developing nations still do not have plans for dealing with a pandemic — and some that do have simply cut and pasted versions of plans from developed countries, which do not apply to poorer nations unable to afford vaccines and antivirals.

People in poor nations already suffer from a higher incidence of conditions, such as malnutrition and HIV, that make them more vulnerable to the new virus. In addition, they are likely to be left out of the global scramble for a vaccine, which has already started as nations such as the United States and Britain rush to tie up vaccine contracts.

On the whole, Oshitani says, "this pandemic came too early. If we had had two more years, we would have been better prepared."

Erika Check Hayden



MARS MISSIONS GET IN LINE

No more special treatment from NASA.

www.nature.com/news

NASA

UK science shuffled again

British prime minister Gordon Brown has scrapped the government department in charge of science and higher education, just two years after it was created.

As part of a cabinet reshuffle on 5 June, responsibilities for research and universities were absorbed into a newly inflated business department that will "build Britain's capabilities to compete in the global economy", says a government statement.

Science-policy experts say that the merger, which brings the United Kingdom's science budget under the ultimate control of business secretary Peter Mandelson, might increase the focus on science as the country tries to haul itself out of economic recession. But scientists might increasingly have to justify the economic benefits of their research.

The new Department for Business, Innovation and Skills sees the end of the Department for Innovation, Universities and Skills (DIUS), a short-lived attempt to bring greater coherence to policies surrounding science and higher education.

Higher education will now be part of a more powerful department, but some worry whether it could become "the handmaiden of business", says Bahram Bekhradnia of the Higher Education Policy Institute in Oxford.

Phil Willis, the Liberal Democrat chairman of the House of Commons committee that oversaw DIUS, criticized the move. "The real casualty of this ill-thought-out reorganization is the nation's strategic science base," he says. "Science needs a stable home at the heart of government policy."

But Martin Rees, president of the Royal Society, praised the decision. "Science and innovation should be the bedrock on which the economy builds as we come out of the current recession," he says. "Placing science alongside business and enterprise should help to make that happen."

Higher-education minister David Lammy is moving to the new department, as is science minister Paul Drayson, who takes on extra responsibilities as defence procurement minister. The former head of DIUS, John Denham, will head the communities and local government department.

The arrangement is likely to be temporary, as the unpopular Labour government must call an election by June 2010.

Richard Van Noorden

H1N1 update

As confirmed by the World Health Organization, as of 8 June:

Cases: 25,288, in 73 countries

Deaths: 139, in 6 countries

AUSTRALIA: Quarantines all players with the Queensland Origin rugby team after a player tests positive for swine flu.

For Nature's swine flu coverage, see www.nature.com/swineflu.

CDC

Moon mission tackles water question

NASA orbiter will hunt for water ice that could be used as a resource by future astronauts.

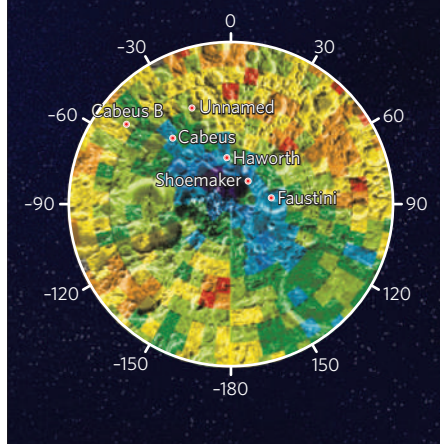
NASA's Lunar Reconnaissance Orbiter (LRO), scheduled for launch on 17 June, should end a long-standing debate on whether the Moon has water ice. For nearly half a century, scientists have argued over the idea that bits of ice hide within the frigid darkness of permanently shadowed craters at the Moon's poles. "I sure hope it settles the debate," says Richard Vondrak, LRO project scientist at the Goddard Space Flight Center in Greenbelt, Maryland.

The spacecraft will also mark a beginning. It is the first in a series of planned NASA lunar missions, the next will launch in 2011, to help scout for the return of astronauts to the Moon. If ice is found, the water could be split into hydrogen and oxygen — for rocket fuel and lungs — to help sustain a lunar outpost.

There are reasons to suspect that ice exists there. A fusillade of water-bearing comets and asteroids could have deposited it, or hydrogen ions carried by the solar wind could have combined with oxygen trapped in minerals on the Moon. Yet there has been no solid detection of ice. The first suggestion came from a radar experiment on the 1994 Clementine mission. The 1998 Lunar Prospector mission added weight to the idea when it found an excess of hydrogen at the poles (see picture, above), which would be consistent with water concentrations of 1% or so. Studies with Earth-based radar, though, have not confirmed the results.

ICE AT THE POLES

The Moon's south polar region may contain ice in permanently shadowed craters. In October, NASA will send a rocket to crash into one of these six target craters in search of ice. Blue represents elevated levels of hydrogen and thus perhaps water.



The case against ice would have been stronger had the Earth observatories been able to peer into the darkest parts of the polar craters, but they could not (see 'Target: Moon'). The first dedicated radar to do so has been that on Chandrayaan-1, India's current lunar orbiter. Stewart Nozette, a Clementine

veteran who is a science team member on the Indian instrument and principal investigator for a similar one on the LRO, says the Chandrayaan team is seeing intriguing signals at the lunar north pole that could be consistent with ice, although the team is still calibrating its data. Together, Chandrayaan and the LRO are scheduled to perform a 'bi-static' observation this summer, with one transmitting and the other receiving, to see how the polar surfaces reflect at different angles — which could also illuminate the ice question.

Neither of the two other recent Asian orbiters had ice as their main target. Kaguya, a Japanese lunar orbiter that will crash into the surface on 10 June, focused on a map of the Moon's gravity field, and Chang'e 1, a Chinese orbiter, mapped the Moon's minerals before its planned crash on 1 March.

Other instruments on the LRO will 'see' into the dark of the crater shadows. A neutron detector will repeat the Lunar Prospector experiment with greater resolution and sensitivity. A temperature mapper will look for hot and cold spots, down to 30 kelvin above absolute zero. And an ultraviolet instrument will look for surface frost using the illumination of starlight.

Six of the LRO's seven instruments will make measurements aimed to detect ice — a surprisingly narrow focus, according to Alan Binder, principal investigator for Lunar Prospector. "It

TARGET: MOON

For all of the Lunar Reconnaissance Orbiter's (LRO) remote sensing, answering the question of whether there is ice on the Moon will probably only be resolved *in situ*. "The only way you're going to know if it's water ice is by sticking your finger in it," says Alan Binder, director of the Lunar Research Institute in Tucson, Arizona.

To this end, the Lunar Crater Observation and Sensing Satellite (LCROSS) is meant to plunge into a shadowy crater near the lunar south pole (see picture, above) in the hopes of kicking up a plume of dust — and, potentially, water vapour — high enough to be seen.

LCROSS is a two-piece experiment: a shepherding spacecraft full of observing

instruments, and the upper stage of the Atlas V rocket on which LCROSS and the LRO will launch. An hour after launch, the LRO will head straight to the Moon, and LCROSS will begin orbits around Earth that put it on a collision course with the Moon for 8 October. The shepherding spacecraft will follow the rocket into the surface, four minutes behind.

There have been other such artificial impacts. Binder, the principal investigator for the 160-kilogram Lunar Prospector, oversaw that mission's intentional crash near the south pole in 1999. The 2,300-kilogram Atlas rocket should make a bigger crater.

In February, LCROSS mission

engineers discovered that valves in the spent upper stage would slowly leak residual fuel, causing wobbles that would need to be counteracted in space with pulses of precious fuel. The team asked for an additional launch delay, from 2 June to 17 June, that gave more time to examine the problem but also, because of the particular orbital mechanics, minimized fuel use. If the problem is worse than expected, mission managers have a contingency plan: they would put LCROSS into an intentional spin that counteracts the wobbles.

There are also uncertainties about the impact itself. The Atlas upper stage will collide at 2.5 kilometres per second, at an angle of 85° — much higher and faster

than previous impacts. Given its tin-can-like structure, the rocket should flatten like a pancake and create a crater about 20 metres wide and 3 metres deep.

Just about everything that can will be watching. Anthony Colaprete, LCROSS project scientist at the Ames Research Center in Moffett Field, California, has lined up observing time at nearly a dozen ground-based observatories, two in Earth orbit and two in lunar orbit. The shepherding spacecraft will be watching, too, as it approaches the plume from above. In its last moments, as it flies into the plume, it will look sideways for watery spectroscopic signatures in the dust-dimmed sunlight.

E.H.

NASA

TOP TARGETS

Although searching for polar ice is a top mission goal, Lunar Reconnaissance Orbiter scientists are also interested in other parts of the Moon. Here are 6 of 50 top-priority sites for the high-resolution camera.

Ina 'D-caldera' In 2006, scientists claimed that gases actively vent from this enigmatic site.

Gruithuisen domes Possible silica-rich volcanoes that resulted in domes, unlike the smooth, pool-like maria.

Rimae Prinz A possible hollow lava tube could offer radiation shielding for an outpost.

Apollo 15 site A chance to observe space weathering of astronaut tracks.

Alphonsus crater Deposits from a pyroclastic vent from a 'fire-fountain' volcano.

Sulpicius Gallus Strange, dark materials could provide a deep window into direct sampling of the Moon's mantle.

E.H.

really is focused on the water question," he says. That may reflect the fact that the \$500-million-plus LRO was paid for by the NASA division that advances human exploration, rather than its science division. Mark Robinson, principal investigator for the spacecraft's main camera, acknowledges that its suite of instruments would probably have been different had its mission arisen via the science directorate.

But Robinson, a geologist at Arizona State University in Tempe, says there will be plenty of data to satisfy scientists of any stripe. The high-resolution camera, working 50 kilometres above the Moon's airless surface, will have a resolution of 1 metre, finally giving scientists maps of the Moon as good as those available for Mars.

The camera will focus first on 50 sites determined by the exploration office (see 'Top targets'), but many of the sites have overlapping science questions. For instance, many of the Moon's maria — sea-like plains of basalt — are rich in titanium, which could be mined. But mapping the deposits could also help scientists understand mysteriously wide differences in titanium concentrations.

Just because NASA's astronaut office is interested in the LRO's data doesn't mean that scientists aren't, says Robinson. "You can't separate the two," he says. "Science enables exploration, exploration enables science."

Eric Hand

Last weather ship faces closure

Leading climate scientists and oceanographers are urging the Norwegian government to revise or postpone the decommissioning of the world's last stationary weather ship.

Located at 66° N in the Norwegian Sea, some 450 kilometres off the coast, the M/V *Polarfront* maintains Station M (Mike), the last of what was a network of 13 weather stations in the North Atlantic. The International Civil Aviation Organization set them up in 1948 to support air traffic, but by 1974 only four were left, and the penultimate station closed in the 1990s.

Now, as satellites, buoys and reports from moving ships have eclipsed the stationary station for weather-forecasting purposes, the Norwegian government says it cannot justify the expense and intends to decommission the ship by 31 December. Annual operational costs are €2.5 million (US\$3.5 million), most of which is paid for by the Norwegian Meteorological Institute in Oslo. The station constitutes more than half of the institute's running expenses for all meteorological observations in the country, including weather radars.

Researchers protest that the shutdown will harm crucial observational programmes in the climate and ocean sciences. "It's a blow," says Ingunn Skjelvan, a chemical oceanographer at the Bjerknes Centre for Climate Research in Bergen, Norway.

"Equally reliable ocean time series are just not available from satellites, buoys or drifting vessels," says Margaret Yelland, a physical oceanographer at the National Oceanography Centre in Southampton, UK.

In more than 60 years of observations, Station Mike has collected measurements of temperature and salinity down to 2,500 metres five times a week — the

world's longest time series of its kind.

Among other things, the station has provided key information on the long-term variability of the Nordic seas and on changes in Atlantic Ocean circulation. "These are not oceanographic curiosities," says Tom Rossby, an oceanographer at the University of Rhode Island in Narragansett. "They are of fundamental importance."

Shutting down the *Polarfront* "will have dramatic negative consequences for climate change research, as irreplaceable long-term measurements will come to an end," wrote Howard Cattle of the World Climate Research Program in a letter to Norwegian research authorities. "Other weather ships were withdrawn from service by the mid-1990s, before the importance of

such sustained observations for understanding climate was fully realized."

Deploying and calibrating hydrographic moorings to

replace the *Polarfront* for climate-quality measuring will require at least a couple of years, experts say.

Norway's research minister, Tora Aasland, has acknowledged the urgent need for new monitoring infrastructure. In a response last month to letters from worried scientists, she said that the Research Council of Norway is prepared to consider scientific proposals for work at Station Mike, and that a decision on funding those could be made this summer.

Even so, the days of the *Polarfront* seem to be numbered. "I have received an overwhelming number of statements stressing the importance of continuing the operation of the ship, but no further offers to share the expenses," says Anton Eliassen, director of the Norwegian Meteorological Institute.

Quirin Schiermeier

"I have received no further offers to share the expenses."



Without funding from elsewhere, the M/V *Polarfront* will be decommissioned.

M. YELLAND



SEX SUPPRESSED IN PLANTS

Pollen and ova produced without meiosis.

www.nature.com/news

J. BURGESS/SPL

SPECIAL REPORT

Quantum dots go large

A small industry could be on the verge of a boom, reports **Katharine Sanderson**.

Nanocrystals called quantum dots have promised to revolutionize display technologies, solar power and biological imaging for more than a decade. Yet the quantum-dot market has remained small, with a handful of companies selling dots directly to researchers, using the particles to develop their own products or licensing their technologies to partners.

"Quantum dots have been around for quite a while, but they're taking a really long time to mature," says David Hwang of the market-analysis company Lux Research in New York. A key barrier is price: quantum dots can cost anywhere from US\$3,000 to \$10,000 per gram, restricting their use to highly specialized applications.

But industry analysts are now predicting extremely rapid growth for the market over the next few years, driven by demand for energy-efficient displays and lighting, and enabled by cheaper, more efficient manufacturing processes. In September 2008, market-research company BCC Research of Wellesley, Massachusetts, predicted that the market for products relying on quantum dots would grow from \$28.6 million in 2008 to \$721 million by 2013, with particularly rapid growth in the optoelectronics sector from 2010 (see graph).

Dots by the kilo

Most commercially available quantum dots have a semiconductor core, often a mixture of cadmium and selenium, measuring about 2–10 nanometres in diameter. This core is surrounded by a shell, usually of another semiconductor material, and an outer polymer or inorganic layer.

The small size of the dots gives them unique properties. Photons hitting a dot excite an electron from the semiconductor material, leaving a positively charged 'hole': this electron-hole pair is called an exciton. In a bulk semiconductor, excitons can have a range of energies within a continuous band. But in the nanoscale dots, excitons occupy distinct, quantized energy states. This means that as each excited electron recombines with a hole, it emits a photon with a specific, predictable wavelength. Smaller dots give off blue light, whereas larger dots of the same material appear red.

Although quantum dots were initially made using expensive techniques borrowed from the computer-chip industry, most dots are currently made by 'wet' chemistry methods, such



There's gold at the end of this rainbow.

as squirting hot solutions of organometallic reagents into a solvent. But scaling up this technique — essential for reducing the cost of quantum dots — has been troublesome, because it is harder to maintain the right temperature in larger volumes.

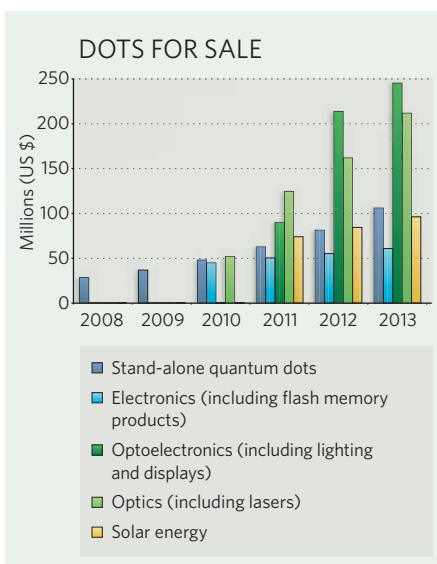
Quantum-dot company Nanoco, based in Manchester, UK, may have an answer. The company, which was spun out of the group of chemist Paul O'Brien at the University of Manchester in 2004, has developed a lower-temperature batch process that uses small seed molecules to grow kilograms of dots from a solution of reagents. And Voxel, based in Beaverton, Oregon, has begun trialling a continuous production process, which can manufacture kilogram quantities a week of most quantum dots for less than \$10 per

gram, according to chief executive George Williams.

The first commercial application of quantum dots took off in 2002, when the Californian nanotech startup Quantum Dot Corporation launched its first quantum dot bioimaging agent. Chemical groups attached to the outside of the dots can hook onto particular cells, for example, giving a useful glowing tag to track the cells' movement. Several different sizes of dots emitting various colours can be used to track multiple cellular processes at the same time.

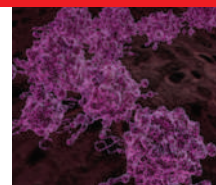
Bioimaging could receive a boost from recent research that offers a solution to a long-standing yet poorly understood problem. Some quantum dots, particularly those prepared by wet chemistry, tend to blink on and off at random. Last month, Todd Krauss at the University of Rochester, New York, and his colleagues announced that they had managed to eliminate blinking by blending the boundary between a core of cadmium, selenium and zinc, and a shell composed of the latter two elements (X. Wang *et al. Nature* 459, 686–689; 2009). Krauss thinks that non-blinking particles may be a decade away from large-scale manufacture, and it is also unclear whether the discovery will apply to other quantum dots. But if it does, "the field will take a fairly substantial leap forward," he says.

Even though bioimaging generated plenty of business for quantum-dot developer Evident based in Troy, New York, the company is branching out into using its dots to fine-tune the colours of light-emitting diodes (LEDs). The dots absorb a relatively wide range of wavelengths produced by the LED, while emitting light in a much narrower range, producing very



X. GAO

SOURCE: BCC RESEARCH



WOMEN LACKING IN TRIALS

Cancer-drug studies don't reflect disease incidence.
www.nature.com/news

PUNCHSTOCK/PURESTOCK

specific colours. The company already sells decorative strings of quantum-dot LEDs in colours from cranberry to tangerine.

Meanwhile, Michael Edelman, Nanoco's chief executive, says that a backlight for flatscreen televisions that relies on the company's dots should be on the market in about 18 months. It will combine red, green and blue quantum-dot LEDs that are more efficient and emit less heat than conventional backlights, he says.

And QD Vision of Watertown, Massachusetts, is about to enter the commercial lighting market. The company was spun out of the Massachusetts Institute of Technology in 2004 by five academics, including electrical engineer Seth Coe-Sullivan, who is now chief technology officer, and Mouni Bawendi, who helped to develop the wet-chemistry method of creating cadmium-based quantum dots (C. B. Murray, D. J. Norris and M. G. Bawendi *J. Am. Chem. Soc.* 115, 8706; 1993).

QD Vision teamed up with Nexxus Lighting in Charlotte, North Carolina, to produce quantum-dot LED lights suitable for commercial and domestic lighting. Unveiled in early May, the lights combine the efficiency of LED

lighting with the warmer colour of incandescent bulbs, says Dan Button, chief executive of QD Vision. He hopes that the lights will provide an energy-efficient alternative to incandescent or halogen lamps, without producing the harsh white light of fluorescent bulbs and conventional LEDs, which often deters consumers. The company expects to begin shipping the lamps, which will cost between \$50 and \$100, by the end of this year.

This kind of collaboration is a common trend in the market, says Hwang. "A lot of these companies are leaning heavily on large corporations to do development and incorporate them into their products," he says.

Shining light

Solar power could also benefit from the tiny dots. Researchers are starting to use them in solar concentrators — flat plates that channel light from a large area and concentrate it onto a solar panel. Organic dyes have already been used in the concentrators to absorb and re-emit light, which then bounces through the plate until it hits a solar cell mounted on the edge.

But dyes are not good at harvesting the red

light in sunlight and are prone to degradation over time. Quantum dots are more robust, can collect light from the far-red and ultraviolet ends of the spectrum, and are also better at capturing the diffuse light of cloudy days. But Amanda Chatten, who is developing solar concentrators at Imperial College London, says that she struggles to buy quantum dots that perform efficiently.

And price is still a significant barrier. "At the moment, quantum dots are about a thousand times more expensive per gram than organic dyes," says Chatten. For a commercially viable solar concentrator, she says that quantum-dot prices will need to drop by a factor of between 100 and 1,000.

Yet despite the global economic downturn, businesses are optimistic. On 1 May, Nanoco started trading on the London Stock Exchange AIM market and has seen its share price rise from £0.21 to £0.56 as of 8 June. Edelman says that the company is busy completing a new, large-scale production facility in Manchester, with a second plant planned in Asia. "If this market is going to take off," he says, "we'll need a lot of material." ■



SLOWING EROSION
Does vegetation stop
coasts washing away?
www.nature.com/news

PUNCHSTOCK/NATIONAL GEOGRAPHIC

Grant applications swamp agency

Peer-review system for National Institutes of Health grants is stretched to its limits.

The US National Institutes of Health, already groaning under the weight of grant applications brought on by a \$10.4-billion economic stimulus package, is likely to be inundated with a second tidal wave of applications this autumn that would send success rates plummeting, agency officials predict.

The flood of applications for Challenge Grants, a new category created by the stimulus funds, means that the success rate will probably be less than 1%. If scientists rejected for Challenge Grants this summer resubmit their applications for standard investigator-initiated 'R01' grants in the autumn, paylines — the percentage of fundable applications — will also bottom out, officials said at an 8 June meeting of the NIH's peer-review advisory committee.

"The NIH budget for 2010, unless there's another miracle, is a very small increase," said Story Landis, director of the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland. "So if in fact all those grants come in, the payline will go to nothing."

"To the low single digits," interjected Antonio Scarpa, director of the agency's Center for Scientific Review. His back-of-the-envelope calculations project that the agency would receive 30,000 R01 proposals this autumn, many of them reincarnated Challenge Grants. Furthermore, he calculated that when the unsuccessful applications from that round are resubmitted yet again, the agency will face 25,000 review decisions in October 2010.

Reviewing at this magnitude "is not sustainable," Scarpa said. The agency is already reviewing more than 40,000 investigator-initiated applications; in a typical June review round, that figure would be 16,000. The NIH normally uses around 8,000 reviewers; it is now up to about 30,000.

More than 21,000 applications have been submitted for Challenge Grants (see table).

Others at the advisory meeting said that the serial wave of applications is being unwisely encouraged by university deans of research. "We have been hearing that deans are telling people: 'If you don't get a Challenge Grant, you turn around and submit another research

THE SURGE

Type of NIH grant	Due date	Size	Number of applications received, as of 4 June	Total funds available initially
Challenge Grant	27 April	\$1 million (\$500,000 per year for two years)	About 21,000	\$200 million
Grand Opportunity Grant	5 June	More than \$500,000 per year for two years; no maximum	2,697	\$200 million by September 2010
Competitive revisions to existing NIH grants	23 April	No limit, but there is a \$100,000 limit on equipment purchases	2,894	\$1 billion
High-end instrumentation grants	6 May	Between \$500,000 and \$8 million	561	\$160 million
Extramural facilities improvement grants	6 May; 17 June; 17 July	Between \$2 million and \$15 million	Applications still open	\$1 billion
\$291 million of the \$800 million allocated to the Office of the Director remains unallocated to any grant mechanism, and must be spent in 2010.				

SOURCE: NIH

project grant.' This is a really serious problem," said Louise Ramm, deputy director of NIH's National Center for Research Resources. "It's frightening."

Garret FitzGerald, chair of pharmacology at the University of Pennsylvania in Philadelphia, added: "Deans have been told they will be rewarded by the number of grants that go out the door. This is irresponsible behaviour."

At universities, "the sponsored-projects offices have been flooded by the number of grants going in. That flood is now becoming a tsunami," Mary-Claire King, a human geneticist at the University of Washington in Seattle, said last week. "They need an early-warning system."

The NIH is in the process of modelling the likely downstream impact of the stimulus spending. That is difficult to gauge at this point, acting NIH director

Raynard Kington said last week, in part because it also depends on how much money Congress gives the agency in 2011 and beyond. Its budget for fiscal year 2009 is \$30.3 billion; last month, President Barack Obama proposed a \$443-million, or 1.4%, increase for fiscal year 2010.

The NIH, which is already in the middle

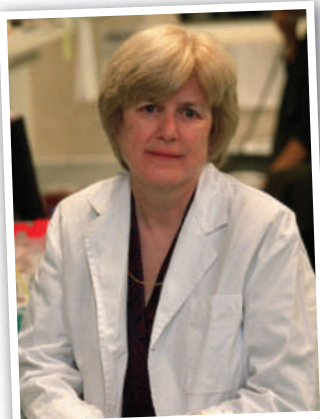
of implementing changes to its peer-review system, has responded to the Challenge Grant flood by recruiting thousands of extra reviewers and adopting 'editorial board' reviews, which involve initial reviews by e-mail and final assessments by boards of highly accomplished scientists.

Among concerns voiced by Kington's advisers at a 4 June meeting is the potential burden imposed by new quarterly reporting requirements for the stimulus money. These have not come into effect yet, but are expected to do so by late summer. The Office of Management and Budget is setting up a reporting system to track the overall \$787 billion in stimulus money; reporting will be consistent across all federal agencies.

"All of us in the extramural world are concerned about what this reporting requirement is going to be and how we are going to manage it," said King. She implored Kington and his staff to adjust the government's template, once finalized, so that "an ordinary mortal scientist" can fill it out.

Kington said it is highly unlikely that the NIH would be able to modify a reporting template signed off by top administration officials as a universal requirement. But, he said, the stimulus money has launched a discussion about how the United States tracks the impact of its science funding. "That's going to be here to stay."

Meredith Wadman



Mary-Claire King: problems ahead.

NEWS.COM



SKB

A mock-up of Sweden's nuclear-waste store.

Sweden finally picks nuclear-waste burial site

After almost two decades of deliberation, Sweden last week announced a final destination for its nuclear waste. The community of Forsmark, around 100 kilometres north of Stockholm, will play host to a long-term storage facility for some 12,000 tonnes of spent nuclear fuel.

The fuel will be housed 500 metres underground in copper canisters buried in bentonite clay. Geologists believe that the combination will keep the fuel geologically stable for up to 100,000 years — the time it will take for the radioactive elements inside to decay to safe levels.

Svensk Kärnbränslehantering, the company tasked with nuclear-waste disposal, hopes to have the SEK12.5-billion (US\$1.6-billion) facility fully operational by 2023.

German science secures historic windfall

German universities and research organizations will receive an €18-billion (US\$25-billion) windfall over the next decade, after Chancellor Angela Merkel and the prime ministers of the country's 16 states agreed on 4 June to finance the package.

In April, the weak economy had made some states waver in their intention to contribute to the proposed programmes, and Peer Steinbrück, Germany's finance minister, had suggested that a final decision might need to be postponed until after September's federal parliamentary elections.

The funding extends three initiatives — intended to strengthen research budgets, encourage university competition and prepare for a rise in student numbers —

which together account for the largest single investment in science and higher education in Germany's post-war history.

For a longer version of this story, see <http://tinyurl.com/l768c6>.

Rotavirus vaccines win global recommendation

The World Health Organization (WHO) has recommended that health authorities in all nations routinely vaccinate young children against rotavirus, which causes 2 million hospitalizations and 500,000 deaths from severe diarrhoeal dehydration every year.

Rotavirus vaccines are already recommended for use in the Americas and Europe. But more than 85% of deaths caused by rotavirus occur in developing countries in Asia and Africa, says the WHO.

The recommendation was made by the WHO's Strategic Advisory Group of Experts, which reviewed a clinical trial of GlaxoSmithKline's Rotarix vaccine in South Africa and Malawi that cut the occurrence of severe diarrhoeal episodes. RotaTeq, a rotavirus vaccine developed by Merck, is being tested in clinical trials in Japan, India, Mali, Ghana, Kenya, Bangladesh and Vietnam.

Californian universities hit by state's budget woes

Californian universities are reeling from unprecedented budget cuts triggered by the state's deficit of more than \$24 billion.

The ten-campus University of California expects to see US\$800 million cut from the \$3.2 billion in state funds it was due to receive over the 2009–10 operating year. And the 23-campus California State

University system, which mainly provides undergraduate degrees, anticipates that at least \$580 million will be sliced from its \$2.7-billion state award.

University of California executives met on 3 June to discuss how they would cope with the cuts in preparation for what is expected to be a turbulent summer. The budgetary axe began taking its toll after voters in last month's state-wide election rejected a series of proposals designed to ease California's deficit.

For a longer version of this story, see <http://tinyurl.com/mwaadb>.

Legal accord eases path to Europe's research facilities

A standard legal framework governing all pan-European research facilities should make it easier for nations to develop large shared projects, European Union (EU) research ministers said after reaching agreement on the scheme on 29 May.

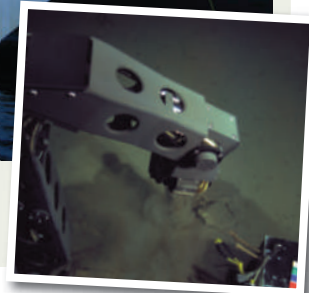
The framework, proposed last year by the European Commission, means that countries will no longer have to debate afresh — under a hotchpotch of national and EU laws — the structure and governance of each European Research Infrastructure (ERI), as collaborations are called. Future ERIs — such as the European Spallation Source, a neutron-science facility (see *Nature* 459, 626; 2009) — will be granted the status of an international organization, and exempted from VAT (value added tax) and excise duty.

"It will bring down barriers to investments in science and research," said Czech education minister Miroslava Kopicová, who chaired the Competitiveness Council meeting in Brussels.

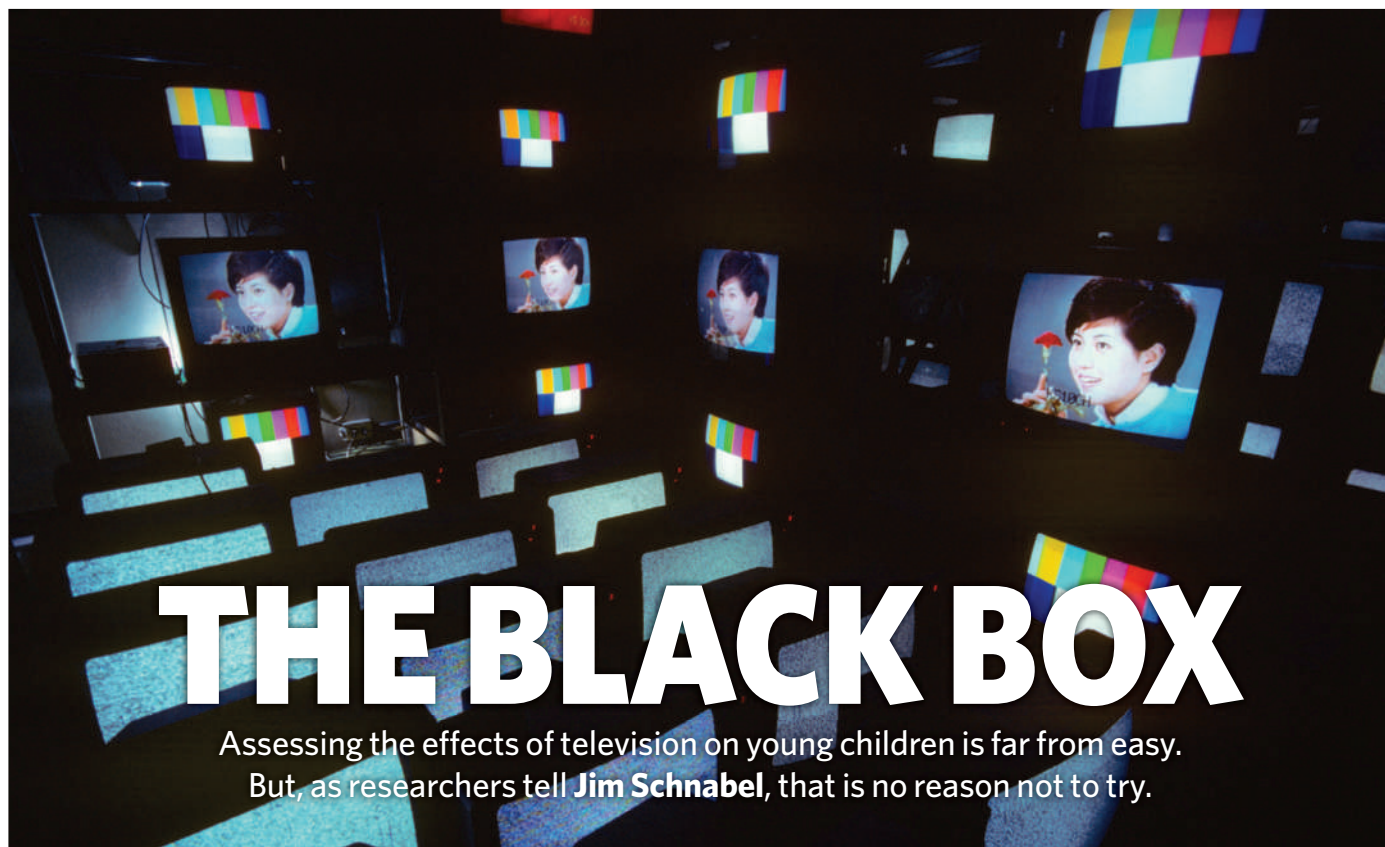
Submersible plumbs the depths

For the first time in more than a decade, scientists have penetrated the deepest parts of the Pacific Ocean. The deep-sea remotely operated vehicle *Nereus* dived 10,902 metres to the bottom of the Challenger Deep in the Mariana Trench on 31 May — the first visit since that of Japan's *Kaiko* submersible in 1998.

Nereus, built by researchers from the Woods Hole Oceanographic Institution in Massachusetts (see *Nature* 437, 612–613; 2005), was dropped from the research vessel *Kilo Moana*, and spent ten hours on the bottom gathering samples and sending back images (pictured, inset).



M. HEINTZ/WOODS HOLE OCEANOGRAPHIC INST.



N. BENN/ALAMY

THE BLACK BOX

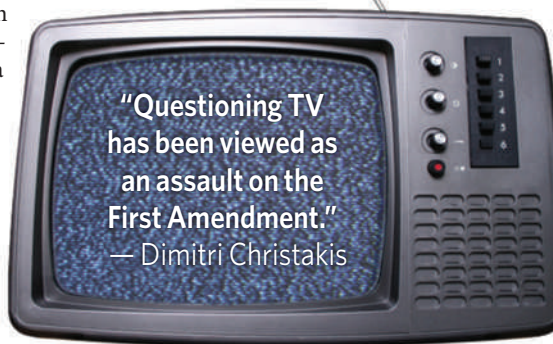
Assessing the effects of television on young children is far from easy. But, as researchers tell **Jim Schnabel**, that is no reason not to try.

In 1998, Dimitri Christakis took time off from work to care for his two-month-old son. At home he found himself watching television to pass the time — “more TV than I had ever watched in my life”, he remembers. Soon he noticed that his infant son was watching too. Even CNN kept the boy glued to the screen. “Obviously he wasn’t following the news,” says Christakis, a professor of paediatrics at the University of Washington in Seattle.

Christakis realized that the jumpy images on the screen were engaging the child’s ‘orienting response’, a basic attentional reflex that directs the senses towards a sudden change in the environment. He wondered about the long-term effect of this on a brain that was at such a sensitive developmental stage. Could it alter the brain to ‘expect’ overstimulation, so that ordinary reality would thereafter seem dull by comparison? And could such a mechanism help to explain the ongoing tsunami of attention deficit hyperactivity disorder (ADHD) diagnoses, whose rise had roughly coincided with the dramatic increase in media consumption in Western societies?

Christakis decided to try to address these questions with research. Together with several colleagues, he examined a database called the National Longitudinal Survey of Youth. After analysing some 1,300 children for whom the appropriate data were available, they found that on average, a child who had watched two hours of television per day before the age of three was 20% more likely to have attentional problems at the age of seven, compared with a child who had watched none.

Christakis and his colleagues published their results in 2004 (ref. 1). Then, working with public-health expert Fred Zimmerman, who is now at the University of California, Los Angeles, Christakis did a follow-up study² with a different longitudinal sample, showing that the link to later attentional problems was particularly strong for cartoons and other entertainment programmes watched before the age of three. For educational programmes, such as the gently paced US series *Mister Rogers’ Neighborhood*, they found no such link.



These studies were among the largest and most persuasive ever to have linked TV to reduced attention and, as such, they made splashes in the media and the research community. But as observational studies, they had their limitations. An association between TV watching and later attention problems did not prove that one had caused the other. A host of other factors, such as the socio-economic

status of a household, could also have contributed to the association, and although researchers typically try to take those other factors into account, they can’t always do so accurately. So, like other scientists who had addressed this issue, Zimmerman and Christakis concluded that more research was needed.

In particular, Christakis believed that what was needed was a large-scale intervention study, a clinical-trial-type experiment in which one randomly selected group of children would be assigned to watch only a small amount of educational TV, whereas the other group would watch whatever their parents normally allowed. However, Christakis’s 2006 proposal — which would have enrolled 900 children, reduced TV exposure in half of them in the first two years of life and assessed attention and related cognitive functions until the age of four — was turned down by reviewers for the US National Institutes of Health (NIH). Even now, the NIH does not have an interventional study under way or planned in this area — nor does Christakis know of one under way or planned anywhere else in the world.

This apparent lack of follow-through seems to be part of a broader phenomenon. On the one hand, there is fairly convincing evidence that some TV content, such as the widely aired *Sesame Street*, can benefit children in a certain age range. On the other, a great deal of research suggests that some TV content can be harmful — yet despite the ominous public-health implications, little seems to be done about it.



"Media are arguably the most ubiquitous environmental influence on kids' health and development," says Michael Rich, a paediatrician and media-effects researcher at Harvard Medical School in Boston, Massachusetts. "The problem is that as a society we have not seen this as enough of a health concern that we've decided to invest in it." Because work on the potential effects of media is seldom distinguished from other kinds of developmental and public-health research, it is difficult to know how much is spent on it. But Christakis estimates that the worldwide total is "below US\$10 million per year".

Tuned out

Some major funding agencies just "don't think that media are as important as other factors in children's lives," says Ellen Wartella, a developmental psychologist at the University of California, Riverside. She notes that at a recent meeting of the Society for Research in Child Development, part of the US National Academy of Sciences, "out of hundreds of sessions on children, only three sessions were devoted to media effects".

Even when media-effects research is done, and its conclusions seem compelling, it appears to have little influence. For example, in at least two recent studies^{3,4}, researchers have tried but failed to find evidence that popular DVDs targeting infants have cognitive benefits, and one of these studies, co-authored by Christakis³, hinted that they might even impair children's language development. Because of research like this, the American Academy of Pediatrics (AAP) discourages parents from letting infants watch TV at all. But apart from France, which last year banned infant-targeted programmes from its broadcast channels, few, if any, countries have policies that restrict infant TV. And Wartella notes that according to surveys, mothers frequently ignore the AAP's advice and don't tell their paediatricians. "They don't want to hear that they shouldn't put their child in front of the screen," she says.

Science's slight impact on media consumption extends beyond issues relating to infants. In the past two decades, researchers have found strong and consistent links between older children's exposure to certain media content and, for example, obesity, eating disorders, aggression, desensitization to violence, sexual promiscuity and the use of alcohol and cigarettes. Brian Primack, a paediatrician at the University of Pittsburgh School of Medicine in Pennsylvania who researches media effects on adolescents, says that for some of these outcomes, "I think we do have enough data [to justify] warning labels".

Studies of adolescents and younger children, meanwhile, show that their media exposure continues to expand rapidly, creating what media-effects researchers have begun to call a

emphasis on media effects amid the complexities of childhood development, says Kevin Durkin, a developmental psychologist at the University of Strathclyde, UK. Among his colleagues, he adds, "there's still a bit of a reluctance to engage with the real world and with things that [children] are actually doing on a daily basis" — and a corresponding preference for theory-driven, laboratory-based research.

Media-effects research, by contrast, has often been considered methodologically messy. Virtually all studies of potential health effects, for example, have been observational studies of real-world populations — populations before and after media exposure, or populations in which different people have different exposures. Because the groups being compared in these cases are not randomly assigned to 'watch' or 'not watch', it is at least conceivable that their differing outcomes are the result of other, perhaps hidden, factors not directly related to media exposure. "There's always going to be that question, because it's not like a randomized clinical trial," says Primack.

As he and other researchers note, the same problem once plagued researchers who wanted to show that cigarette smoking harmed people's lungs. A 'clinical trial' of smoking's health effects, in which groups of people were randomized to smoke or not smoke, would have been grossly unethical. Over time, however, researchers were able to build a case against tobacco with well-designed observational studies, for example showing a 'dose-response effect' in which heavier smoking was linked to higher cancer risk.

But although cigarette smoke is a relatively simple kind of exposure, media exposure is much harder to track. Typically, researchers have lacked the means to monitor precisely what their subjects are watching from day to day for study periods that may last months or even years. They have often relied on self-reporting or parental reports, both of which are considered unreliable. "If you ask the parents about how much TV their kid watches, they tend to overestimate, whereas the kid tends to



Dimitri Christakis assesses child development and media.

'digital childhood'. Is this happening because the evidence for media's harmful effects remains sparse? Or have researchers such as Christakis found themselves up against cultural forces that they cannot defeat with evidence alone?

Researchers argue that a host of factors, both scientific and societal, are to blame. For one thing, researchers such as Christakis, Rich and Primack are paediatricians, yet their proposals for epidemiological-type research on the effects of childhood media are usually reviewed by developmental psychologists. Developmental research traditionally hasn't placed great

SEATTLE CHILDREN'S RES. INST.



Anytime, anywhere: exposure to TV is hard to measure accurately.

underestimate," says Rich. "One way or another the report tends to be biased."

In the past few years, Rich, Primack and others have been experimenting with other monitoring methods such as ecological momentary assessment, in which subjects or their parents report real-time data of their media exposures to researchers using cellphones, personal organizers and even camcorders.

Information overload

On the whole, though, technological evolution may be making it harder to study media's effects. "It's no longer the box in the middle of the living room that everybody gathers around," notes Rich. "We have screens everywhere — in our pockets, you know? I was in a hotel recently and I realized that from where I sat I could see seven screens with seven different things on them. And I wasn't even actively watching media." Similarly, notes Susan Newcomer, a programme officer who oversees some media-effects research at the National Institute of Child Health and Human Development (NICHD) in Bethesda, Maryland, "How do you measure screen exposure if the kid is clicking through 72 web pages a minute? And not just clicking through the web but also texting a friend, listening to music, all at the same time?"

Compounding this problem is the fact that even for a single, uncomplicated medium — such as one television channel — a viewer faces multiple potential influences that may be difficult, if not impossible, to tease apart. There is the stream of content within a television programme, for example, and then there are the contents of interleaved advertisements. Viewers may be affected, too, by the 'formal', non-conceptual properties of TV watching, such as the rapid image-shifts meant to trigger viewers' orienting responses. There is also the likelihood that in watching TV, a viewer will be inert on a sofa instead of exercising, sitting alone instead of socializing, and staying awake instead of sleeping — and all of these behavioural displacements have their own potential health effects.

"You're about to go to bed and you turn on the television to wind down, but then three hours have passed and you've seen *Dracula* again," says Primack, pointing out that a lack of sleep is now believed by some to contribute to obesity as well as mental-health problems.

According to James Griffin, a programme officer whose NICHD branch oversees research on infants and young children, the institute now generally prefers studies that can identify specific cognitive mechanisms of a child's interaction with media. "We are focusing more on looking at what infants and toddlers are capable of learning [from media]," he says.

In theory, interventional studies would get around many of these problems because the

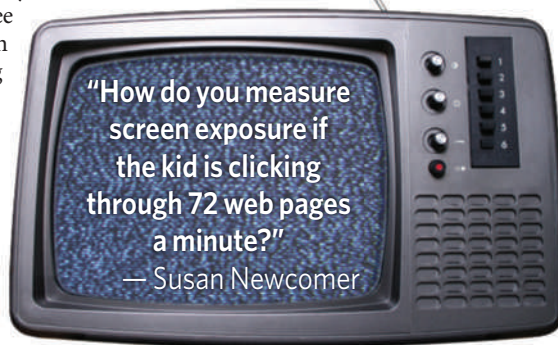
The early developmental stage of these children also means that TV's effects may show up more quickly, thus requiring shorter studies. Without such investigations, Christakis says, "it's going to be very difficult for us to satisfactorily prove to the critics and the cynics that a particular content is harmful".

Media-effects researchers are not just fighting resistance in the scientific community. They say they face a broad cultural resistance too, one rooted in core Western beliefs — that humans have 'free will'; that they choose autonomously and rationally from the information put before them; that a free market in information is better than a restricted one.

"In the United States, questioning any aspect of television has been viewed as a direct assault on the First Amendment," says Christakis. "Tobacco researchers didn't have to deal with that issue." And the issue isn't confined to the United States. In 2005, India's government banned film images of actors using cigarettes — a known motivator for adolescents to start smoking — yet the country's High Court reversed the ban earlier this year, on free-speech grounds.

Researchers say that they confront a related problem called the 'third-person effect'. "Most people will acknowledge that TV is on balance a bad thing; they just don't think it's bad for *their* kids," explains Christakis. "Somehow they think that their kids are immune, or that the way they use TV is different." But the conceit isn't justified, Rich says, because "epidemiologic data [showing a harmful influence] are generated on our kids, not on somebody else's kids".

Media-effects researchers also say that their arguments are apt to be treated, by the public and even by other scientists, as old-fashioned arguments about the immorality of popular-media content, now dressed up as modern health issues. Yet Rich insists that the difference between moral issues and health issues is real and may be crucial in convincing people to change their media-consumption habits. "The



source and content of media could be controlled and monitored. So why not do them? "Good luck getting that past either an NIH review panel or an institutional review board," says Newcomer. Over the time needed for any effect to appear, she explains, it would be "both ethically and pragmatically a real challenge to constrain people not to look at television, for example, or to look at only certain kinds of television".

Yet Christakis argues that such intervention studies are still feasible, particularly when it comes to the effects of TV on very young children. In principle, for such children, TV watching can be more easily controlled by parents, and more easily monitored by researchers.

FROM LEFT: CARO/ALAMY; J. GREENBERG/ALAMY; T. KRUESSELMANIN/ZEFA/CORBIS; PICTURE PARTNERS/ALAMY; S. RAUSSEY/TAXI/GETTY; VARIO IMAGES/ALAMY; IMAGEBROKER/ALAMY



moral issue has helped to stalemate this," he says. "People have different value systems. But if you present them clear data that show what the effect of this media content is on that health outcome, you can get them to agree."

Even if people can agree such things on a rational level, in practice their media-consumption behaviours are deeply entrenched, as even Rich admits. "Parents do tend to say I'm sorry I need to put the kids in front of the tube so I can get the house vacuumed."

Sometimes, governments step in to break cultural stalemates, and in the United States — where the majority of media-effects research takes place — that almost happened. In 2004, Senator Joseph Lieberman (Democrat, Connecticut) introduced a bill called the Children and Media Research Advancement Act (CAMRA). Designed to remedy "the paucity of research" in this area, and enthusiastically backed by researchers including Rich and Wartella, it would have dedicated an NIH budget for media-effects studies during 2005–09, starting at \$10 million per year and ending at \$25 million per year — figures that would have dramatically increased the activity in this field.

The bill never made it past the committee stage. Indeed, a group called Citizens Against Government Waste criticized Lieberman, stating in a press release that his proposed effort "belittles the ability of parents to use common sense in deciding what entertainment is appropriate for their own child's consumption".

CAMRA has been reintroduced several times since then, with funding levels left unspecified. But it still has not made it into law. According to a Senate staffer who didn't want her name used, some senators have objected to the legislative determination of where NIH funds should be applied — which they say the agency has had enough of already. (A 2006 version would have placed the programme with the Centers for Disease Control and Prevention in Atlanta, Georgia.) Other legislators simply have not wanted to spend the money, although on an annual basis

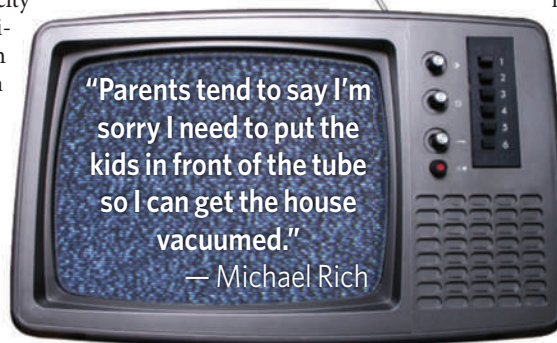
that money would have amounted only to the cost of a few episodes of a hit TV show.

"That in a nutshell is our societal ambivalence towards this subject," says Rich.

Accentuate the positive

Last year, Christakis was able to get a media-effects study funded by the NICHD. Just starting now, the project will look at the effects on preschool children of changing TV content rather than eliminating it. "We're trying to get them to watch less aggression and more prosocial programming," says Christakis. "But we're not telling them to watch less; we just want them to watch better TV." The pro-TV message, he adds, is "something that everybody gets behind".

Rich, too, notes that "the problem with coming at it from a totally negative standpoint is that



people just check out". He urges a greatly expanded research base "that tells us how to live with media, because we're going to have to one way or another". He also foresees that media use will change, albeit 'generationally', as it did with cigarettes. As evidence accumulates that TV can cause health or behavioural problems, he predicts, certain kinds of media use may start to be seen by youngsters and their parents not as forbidden pleasures but as "just dumb".

Christakis is less sanguine. He stopped his own kids from watching TV until the age of two and now places modest restrictions on their Internet use. He worries that, in addition to its

functional entrenchment in people's daily lives, a lot of media content might be physiologically addictive — perhaps less so than nicotine-laden cigarettes, but with a much larger susceptible population. "I think there's a biochemical response to that kind of stimulation, that instant gratification, that rush, which isn't very well understood," he says. "And I think there's a sense in which some people are very reluctant to acknowledge that there's a problem, because they also suffer from it."

Although the concept of behavioural as opposed to drug-induced addiction is still somewhat controversial in Western countries, it is less so in South Korea and Japan. The exposure to electronic media, especially the Internet, seems to be greater there, and the reported prevalence of related addiction behaviours is remarkably high⁵. Christakis suspects that those societies may be "the canary in the coal mine" when it comes to media addiction.

One advantage researchers had in studying the effects of cigarette smoking was that, even if they couldn't experiment directly on humans, they could do so on animals. Christakis now hopes to remove that advantage. This summer he will begin tests on a rodent model of media exposure, comparing rats raised in a 'hyperstimulating' environment of fast-changing sounds and lights, with those raised in a quiet setting, and looking at outcomes relating to attention and addiction. "Animal studies have their limitations," he says. "But they have their strengths too, because you can totally control the animal. You can drill down, so to speak, to a level you could never do in humans."

Jim Schnabel is a freelance writer based in Maryland.

1. Christakis, D. A., Zimmerman, F. J., DiGiuseppe, D. L. & McCarty, C. A. *Pediatrics* **113**, 708–713 (2004).
2. Zimmerman, F. J. & Christakis, D. A. *Pediatrics* **120**, 986–992 (2007).
3. Zimmerman, F. J., Christakis, D. A. & Meltzoff, A. N. *J. Pediatr.* **151**, 364–368 (2007).
4. Robb, M. B., Richert, R. A. & Wartella, E. A. *Br. J. Dev. Psychol.* **27**, 27–45 (2009).
5. Block, J. J. *Am. J. Psychiatry* **165**, 306–307 (2008).

FROM LEFT: APPLY PICTURES/ALAMY; E. KASHI/CORBIS; VARIO IMAGES/ALAMY

The boundaries of biology reach farther below Earth's surface than scientists had thought possible. **Amanda Leigh Mascarelli** delves into how microbes survive deep underground.

LOW LIFE

In February, a team of American and German oceanographers set out on a ship for a little-known destination in the middle of the Atlantic Ocean called North Pond. This patch of sea floor lies on the western flank of the Mid-Atlantic Ridge — the longest mountain range in the world — where the topography of the ocean bottom drops to form a 10-kilometre-long basin rimmed by underwater peaks.

For two weeks, Katrina Edwards, a geomicrobiologist from the University of Southern California in Los Angeles, and her team explored North Pond, collecting samples of the muddy sediments that fill the basin. From their ship, they dropped hollow coring tubes down through 4.5 kilometres of water and into the bottom muck. On lucky days, the equipment went straight through the sediment and struck the underlying rock, which bent the coring barrel into the shape of a banana. Although the collisions sacrificed a few pieces of pipe, they also yielded samples of the delicate interface between the rock and the sediment, one of the targets high on the researchers' wish list.

Edwards had come 7,000 kilometres to look for 'intra-terrestrials' — the microbes inside the sediments and the rocks beneath, where not long ago it was thought that life could not exist. She is among a group of scientists who are learning just how resilient and pervasive life is in the deep earth, both under the sea floor and inside the continental crust. Nicknamed the 'iron maiden' by her colleagues, Edwards is particularly interested in those life forms that feast on iron and that colonize some of Earth's most inhospitable terrain: the igneous crust that reaches to some 500 metres below the ocean bottom. "What I study is essentially the tooth decay of the solid Earth, the microbes that inhabit the nooks and crannies of Earth's molars that are exposed at the bottom of the ocean," says Edwards.

Such areas were largely inaccessible until the 1990s, when new techniques made it possible for scientists to make direct observations of this deep biosphere. In particular, oceanographers have developed sub-sea-floor laboratories known as circulation obviator retrofit kits (CORKs), which seal scientific instruments inside deeply drilled boreholes

and make real-time measurements of life in the deepest, darkest realms of the marine subsurface. To date, researchers have mounted only one scientific drilling mission, in 2002, that was wholly dedicated to this biosphere, but they are poised to launch four more by 2013 through the international Integrated Ocean Drilling Program. "We're right at the cusp of this major breakthrough," says Edwards, who plans to return to North Pond in a year or two.

The North Pond study and others around the world are changing the way scientists think about the deep biosphere. Ten years ago, such low-life microbes were largely regarded as curiosities that represented one of the last frontiers on Earth. Now, scientists have come to appreciate these organisms as integral players in global cycles, helping to replenish key minerals in the ocean and even mediating the climate. "As the science matures, there is an ongoing sense of wonder about what's down there, but we're also coming to understand how they are involved in the biogeochemical cycling and the health of our planet," says Rick Colwell, a geomicrobiologist from Oregon State University in Corvallis (see 'Mining value from deep life', overleaf).

New findings are also leading to insights about the origins of life on Earth and how life might exist on other planets. Although the microbes turn up nearly everywhere that scientists search for them, they often seem to subsist at the very brink of survival, metabolizing so slowly that it has prompted fresh ideas about the limits of life.

Deep-sea sandwich

In 1955, Claude ZoBell, considered to be the father of marine microbiology, probed beneath the sea floor and found microbes there, decreasing in numbers down to a depth of about 8 metres¹. At that time, researchers thought that life would peter out at some point not far below the seabed. Then in the late 1960s, an inadvertent experiment supported the notion of a depauperate deep sea, when the research submersible *Alvin* sank more than 1,500 metres after a cable snapped. The crew of three escaped safely through the hatch, but their lunches were left behind.



When the vessel was recovered 10 months later, the crew was surprised to find their stranded, soaked bologna sausage sandwiches and apples in nearly pristine condition, showing no sign of microbial decay. "This was the popular vision of the deep sea, being too extreme even for significant bacterial life," says John Parkes, a geomicrobiologist at Cardiff University, UK.

The notion of the deep sea as an uninhabitable desert persisted for decades, colouring thinking about the sea floor and what lay beneath it. Then, in the 1980s and 1990s, some of the first missions of the Ocean Drilling Program made it possible for researchers to dig deeper than ever before. When Parkes and his colleagues tried in 1990 to publish results in *Nature* showing that bacteria could colonize much greater subsurface depths than previously thought, they were met with "very sceptical reviews" and the paper was rejected, he says. But in 1994, they succeeded in publishing their results and reported viable microbes living in ocean sediments at depths greater than 500 metres below the seabed².

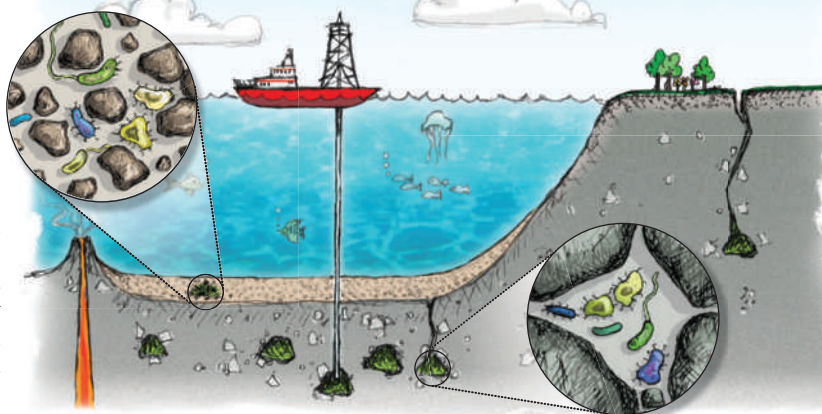
Active bugs

That and subsequent studies showed that microbes could be cultured from samples obtained far below the sea floor. But the techniques at the time could not definitively show that the organisms were alive and actively metabolizing at such remarkable depths, leaving open the possibility that the deep bacteria were dormant, barely living. But in 2005, researchers led by Axel Schippers of the Federal Institute for Geosciences and Natural Resources in Hannover, Germany, showed the presence of intact membranes and ribosomes³ — the first conclusive evidence that bacteria are thriving in 16-million-year-old sediments more than 400 metres deep. Last year, metabolically active microbes were reported in 111-million-year-old sediments buried as deep as 1.6 kilometres below the seabed⁴.

So little is known about microbes that dwell in the deep that scientists have a hard time estimating what fraction of life they represent. A decade ago, estimates derived from work by William Whitman of the University of Georgia in Athens and his colleagues⁵ suggested that one-third of all life on Earth lives in the sub-sea-floor sediments. But most of the samples of microbes from deep sediments have been collected close to shore, meaning that much of the ocean has been underrepresented. Steven D'Hondt, an oceanographer at the University of Rhode Island in Narragansett, recently sampled sub-sea-floor sediments in the North and South Pacific oceans. Those findings suggest that global cell abundances may be an order of magnitude lower than previous estimates, D'Hondt and his colleagues reported last December at a meeting of the American Geophysical Union in San Francisco. Yet Whitman's numbers did not include microbes living in the ocean crust, which

BIOLOGY'S DEEP DARK SECRETS

Subsurface microbial communities have been found within sub-sea-floor sediments, in the underlying ocean crust and far below the surface of the continents.



N. SPENCER

would add to the estimated cell counts, says Edwards.

These hidden microbes are turning up in other unexpected places. In the late 1990s, researchers plumbing the depths of the continental crust in a South African gold mine discovered microbes living at about 3 kilometres below the surface. Plans are now under way to begin drilling in the deepest mine in North America: Homestake in South Dakota, which is to house the US Deep Underground Science and Engineering Laboratory. The mine reaches down nearly 2.5 kilometres, and researchers hope to drill from that depth into rock with temperatures exceeding 120 °C. "The deepest extent of the biosphere is currently unknown," says Tom Kieft, an environmental microbiologist from the New Mexico Institute of Mining and Technology in Socorro. "If we drill deeply enough, we'll reach beyond the upper temperature limit for life, which is thought to be around 121 °C."

Organisms that live in the deep biosphere bear little resemblance to surface bacteria such as *Escherichia coli*, which can be easily cultured in the lab and divide every few minutes. In the sub-sea floor, bacteria and another group of microbes called archaea are slow by comparison, says Bo Barker Jørgensen, a biogeochemist at Aarhus University in Denmark. For organisms buried beneath the surface of the continents, the first estimates suggested that they reproduce on a timescale measured in centuries⁶. And Tullis Onstott, a geomicrobiologist at Princeton University in New Jersey who pioneered much of the exploration for deep terrestrial life in South African gold mines in the late 1990s, estimates that subsurface microbes there may reproduce once every 1,000 years.

Energy crisis

Even with such low metabolic rates, it remains unclear how such organisms sustain themselves. "When we do the calculations, there's not enough energy down there at all for these organisms," says Parkes. "They should all be dead."

The sparse food present in the deep-sea sediments comes from the sunlit layers. There, photosynthesizing plants and algae digest organic matter that eventually rains down in the form of dead algal cells, faecal matter and marine detritus. It settles on the sea floor and accumulates in the sediments over millions of years. Scientists estimate that the majority

Katrina Edwards samples rock from the Atlantic seabed.



J. SYLVAN

of sub-seabed microbial communities graze on this deeply buried organic carbon, contributing to the 'deep carbon cycle'. Only those microbes beneath the sea floor can metabolize these gritty organic leftovers.

Like earthworms that plough the soil and recycle minerals and nutrients, these sub-sea-floor microbes produce carbon dioxide and methane and liberate key elements including nitrogen, sulphur and phosphorus from the sediments. And as fluids circulate through the crust, they carry microbes that can erode the rock, releasing iron and other elements. The circulating fluids take these nutrients back up into the ocean, where they can feed the growth of new biomass. "The carbon cycle of the oceans and of planet Earth reaches deeply into the subsurface biosphere and cannot be understood without the subsurface contribution," says Andreas Teske, a microbiologist at the University of North Carolina in Chapel Hill.

The discoveries of the past decade back up a portion of the 'deep hot biosphere' hypothesis, proposed in 1992 by the late Thomas Gold, an astronomer at Cornell University in New York. In a famous paper⁷, Gold argued that the subsurface supports a mass and volume of life rivaling that present on the surface. But Gold went further to speculate that the deep biosphere subsists on hydrocarbons rising from Earth's mantle, a vast energy source that continually refills oil deposits. That suggestion is no more accepted now than it was at the time.

Climate connection

Even as they help recycle nutrients, the microbes below the sea floor may also have an effect on the planet's climate. Archaea called methanogens produce methane as a by-product of their metabolism. Colwell is working to quantify the rates of that methane production and says that they sometimes fall below the detection limit⁸. Yet the methane builds up over geological timescales and contributes to the formation of 'hydrates', icy cages of water molecules surrounding methane that become wedged in marine sediments. Most methane hydrates are thought to be generated by microbial processes⁹.

Such deposits have apparently become destabilized at several times in Earth's past, releasing enough methane — a greenhouse gas much more potent than CO₂ — to warm the planet substantially¹⁰. As the globe heats up because of human pollution, scientists are becoming increasingly concerned about the potential breakdown of methane hydrates

that are trapped inside thawing Arctic tundra and shallow marine sediments.

Whereas the vast majority of microbes in the deep biosphere rely on leftover organic matter for food, others seem to be getting sustenance from an inorganic source. D'Hondt is investigating whether sub-sea-floor microbes may be obtaining energy from hydrogen that is produced when the radioactive decay of naturally occurring elements such as uranium, thorium and potassium splits water molecules into hydrogen and oxygen. The same process would have occurred on the early Earth more than 4 billion years ago, "and it should be occurring on Mars today," says D'Hondt. "We've found some evidence that as much energy is entering these [sub-sea-floor] ecosystems from radioactive splitting of water as from burial of organic matter."

As yet, though, the identity of any microbes making a living in this way is unclear, he says.

In the terrestrial biosphere, microbes found in a South African gold mine nearly 3 kilometres underground have a similar form of metabolism. They subsist on geologically produced sulphate and hydrogen, free from any dependence on energy derived from the Sun¹¹.

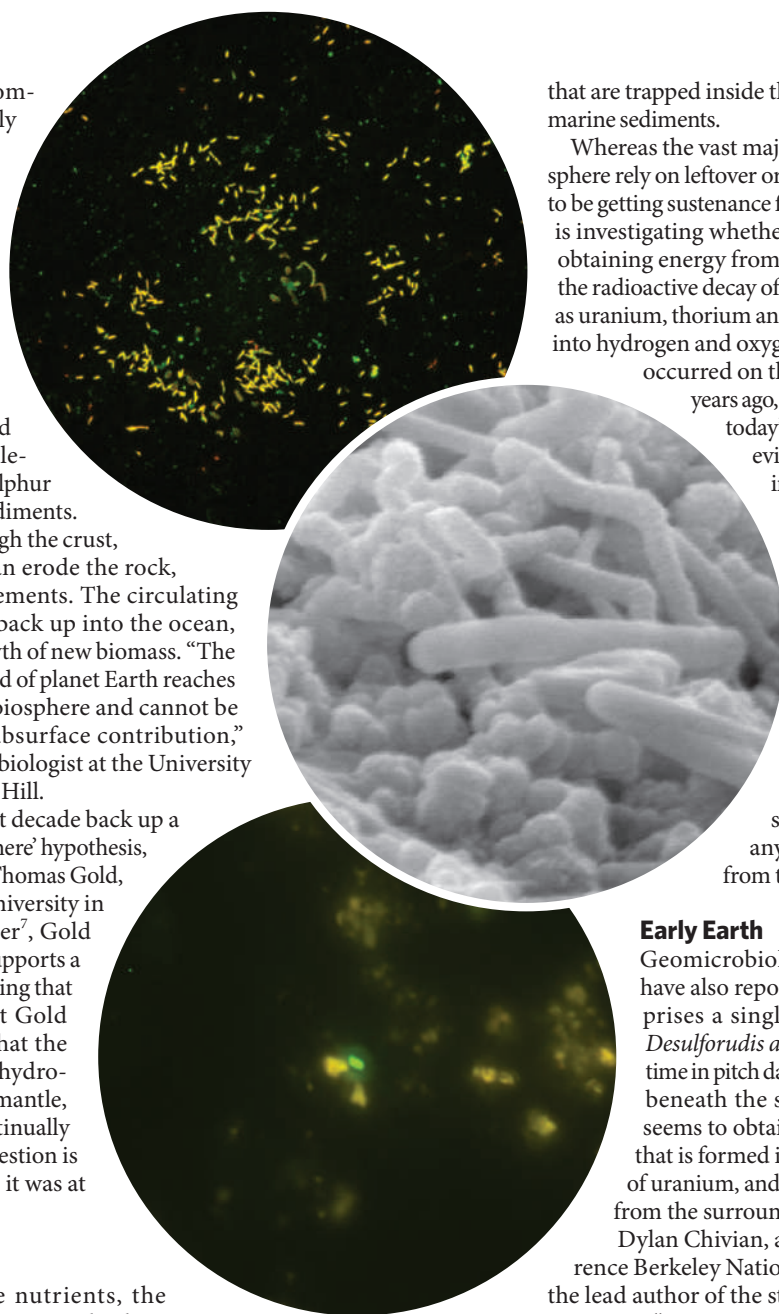
Early Earth

Geomicrobiologists working in South Africa have also reported the first ecosystem that comprises a single bacterial species, *Candidatus Desulforudis audaxviator*, which lives out its lifetime in pitch darkness at 60°C, some 2.8 kilometres beneath the surface of Earth¹². This microbe seems to obtain its energy by reducing sulphate that is formed indirectly by the radioactive decay of uranium, and it can extract carbon and nitrogen from the surrounding rocks.

Dylan Chivian, a computational biologist at Lawrence Berkeley National Laboratory in California and the lead author of the study, says that this single-species ecosystem "points to a mode of life that potentially is what early Earth might have been like", before the atmosphere held much oxygen.

The deep microbes discovered in mines are sometimes found in pockets of water that have been secluded for millions of years, practically making the organisms there living fossils. Fractures in the rocks periodically open and close due to tectonic shifts, locking the water and the microbes into what Barbara Sherwood Lollar, a geochemist at the University of Toronto in Ontario, calls a "series of time capsules". "In these hydrogeologically isolated systems of very great age, it raises all kinds of fascinating questions about how long the microbes have been there, how they've evolved and what that means for our understanding of the origins of life on the planet," she says.

Because microbes that colonize the deep biosphere have mastered the art of living on the margin, the astrobiology

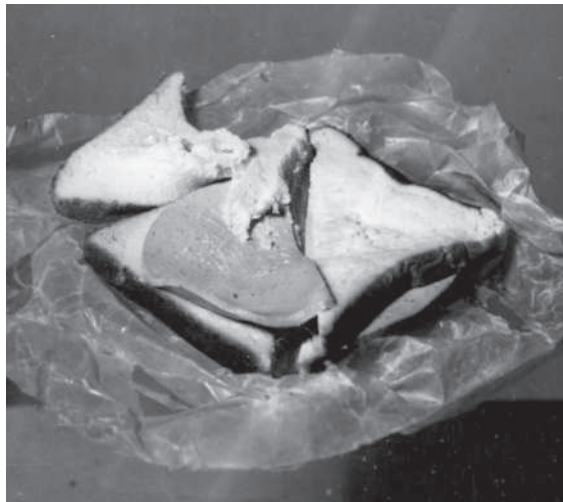


Deep denizens: photomicrographs of microbes from below the sea floor in the Pacific Ocean (top), from the continental crust in South Africa (middle) and beneath the Arctic seabed (green dot, bottom).

B. ORCUTT, K. EDWARDS

M. DAVIDSON, G. SOUTHAM

D.C. SMITH



This sandwich spent months at the bottom of the ocean.

community has also taken a keen interest in them. “Energy limitation is almost certainly a fact of life in extraterrestrial habitats,” says Teske. The anaerobic metabolic reactions that deep-dwelling microbes use are the very chemical reactions that are most likely to support life on other planets, he adds.

“It’s so fascinating that all this microbiology and these processes that were once considered to be in the realm of geology are really supporting a major part of life on Earth,” says Jørgensen. “It’s like discovering a new continent.”

That sense of potential is what motivated Edwards and her colleagues during their recent trip to North Pond. In the waning days of the expedition, the team kept a frenzied pace collecting mud cores and storing them in refrigerators for the journey home. This cruise was a reconnaissance mission of sorts, which will set the stage for future expeditions. In 2010 or 2011, Edwards plans to return aboard the drill ship *JOIDES Resolution* that can penetrate rock and bore down some 500 metres into the crust. Once there, they plan to install a CORK observatory that will peer inside the intra-terrestrial underworld and monitor that environment for a decade. These missions and others, Edwards hopes, will finally shed light on the deep dark biosphere and the lower limits of life itself. ■

Amanda Leigh Mascarelli is a freelance science writer based in Denver, Colorado.

1. ZoBell, C. E. & Morita, R. Y. *Deep-Sea Res.* **3**, 66–73 (1955).
2. Parkes, R. J. *et al. Nature* **371**, 410–413 (1994).
3. Schippers, A. *et al. Nature* **433**, 861–864 (2005).
4. Roussel, E. G. *et al. Science* **320**, 1046 (2008).
5. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583 (1998).
6. Phelps, T. J., Murphy, E. M., Pfiffner, S. M. & White, D. C. *Microb. Ecol.* **28**, 335–349 (1994).
7. Gold, T. *Proc. Natl Acad. Sci. USA* **89**, 6045–6049 (1992).
8. Colwell, F. S. *et al. Appl. Environ. Microbiol.* **74**, 3444–3452 (2008).
9. Kvenvolden, K. A. & Lorenson, T. D. *Geophys. Mono.* **124**, 3–18 (2001).
10. Kennett, J. P., Cannariato, K. G., Hendy, I. L. & Behl, R. J. *Methane Hydrates in Quaternary Climate Change: The Clathrate Gun Hypothesis* (American Geophysical Union, 2002).
11. Lin, L. *et al. Science* **314**, 479–482 (2006).
12. Chivian, D. *et al. Science* **322**, 275–278 (2008).
13. Wymore, R. A. *et al. Bioremed. J.* **11**, 125–139 (2007).
14. Inagaki, F. *et al. Proc. Natl Acad. Sci. USA* **103**, 14164–14169 (2006).

Mining value from deep life

Microbes living far below Earth’s surface have some unusual traits that humans are hoping to exploit. The government of South Africa, for example, launched a programme in 2007 to explore the potential industrial applications of microorganisms and products from the deep biosphere. It is currently evaluating four products for market, says Esta van Heerden, a biochemist at the University of the Free State in Bloemfontein, South Africa.

Enzymes, antimicrobials and antivirals isolated from sub-surface microbes hold promise for the biotechnology industry. One of these enzymes naturally detoxifies an environmental carcinogen from mining ore called hexavalent chromium. In addition, some microbes produce enzymes that can deposit metals such as gold, silver and platinum as nanoparticles. These nanoparticles have unique optical and magnetic properties that can be useful for drug delivery and other applications, says van Heerden.

Rick Colwell, a geomicrobiologist at Oregon State University in Corvallis and his colleagues are investigating the potential for microbes from the deep biosphere to mop up environmental contaminants such as trichloroethene, a chlorinated solvent that leaked

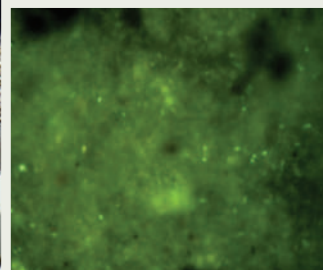
into an aquifer in southeast Idaho as a result of industrial disposal. A community of microbes naturally present in the basaltic aquifer there uses methane in the subsurface as an energy source and co-metabolizes the contaminant, thereby slowly cleaning the water¹³. Colwell expects that these microbes, and other naturally occurring communities in the aquifer that metabolize organic compounds, could reduce clean-up costs by some US\$7 million over coming decades.

Some deep microbial communities might even help humans tackle the problem of climate change. In a volcanically active region near Taiwan, a peculiar lake of liquid carbon dioxide exists under a bed of deep-sea sediments as a result of the extreme pressures 1,380 metres below the sea surface. Fumio Inagaki of the Japan Agency for Marine-Earth Science and Technology in Kochi and his colleagues have found a dominant community of anaerobic microbes that metabolize methane and sulphur and assimilate the CO₂ from the surroundings¹⁴. This community is a perfect natural laboratory to study the potential for disposing of CO₂ in the sub-sea floor, Inagaki says.

Colwell and his colleague Martin Fisk, a marine geologist, are investigating microbial communities that exist in deep volcanic rocks in eastern Washington state where CO₂ will be injected as part of a pilot-scale sequestration project. Colwell wants to know how the microbial communities will respond to the injected CO₂ and whether they can be used to help monitor and verify the presence of CO₂ for the purpose of a carbon-trading scheme in the future. **A.L.M.**



Rick Colwell (above) collecting detoxifying microbes (right) in an aquifer in Idaho.



CORRESPONDENCE

Gene data for endangered species have limitations

SIR — Your News story ‘Time to sequence the “red and the dead”’ (*Nature* **458**, 812–813; 2009) reports on plans to sequence the genomes of endangered and extinct species. Supporters claim that these sequences could help us learn why some species became extinct and provide a scientific argument to warn politicians and the public about which species are endangered, resulting in improved policies. Unfortunately, their claim has several flaws.

Conservation biologists studying a change in genetic diversity over time need many specimens to understand the process of extinction, as you imply. However, a limited set of markers would be adequate for this undertaking, so there is no compelling scientific reason for collecting hundreds of complete genomes.

Although DNA samples may be useful in solving some extinctions attributable to pathogens, genome data would not have helped the now-extinct Chinese river dolphin (*Lipotes vexillifer*), which needed improved habitat — better water quality and a reduction in fishing and boat traffic. Nor would they have helped the northern white rhinoceros (*Ceratotherium simum cottoni*), which, now probably already extinct in the wild, needed better protection from poaching.

The wrong policy decisions could be made on the basis of population genetic data if these are ambiguous or uninformative about a species’ prospects. Take the case of the mammoth and the bison: based on no evidence for a change in mammoth population size over time and the steep decline in bison over some 20,000 years, it might be inferred that bison was at risk of extinction whereas mammoth was not; the reality was the reverse.

For the global preservation of species, it is much more important to reach a timely understanding of ecological requirements and

the effects of invasive species and climate, for example, as well as population parameters that influence responses to harvesting. Otherwise, proposals to sequence “the red and the dead” may quickly be reduced to sequencing just “the dead”.

Michael Hofreiter Evolutionary Biology and Ecology, Department of Biology, University of York, York YO10 5YW, UK
e-mail: msh503@york.ac.uk

How science upholds civilization, human rights and democracy

SIR — I would like to add two important points to the discussions of C. P. Snow’s ‘two cultures’ (*Nature* **459**, 10 and 32–39; 2009).

First, it is not ignorance of the second law of thermodynamics that is the most serious gap in the education of many non-scientists, but a lack of basic understanding of how science works. For example, a recent UK survey, *Public Attitudes to Science* 2008, by Research Councils UK (<http://tinyurl.com/o96lwg>) showed that, although the public overwhelmingly believes that science makes the world a better place, a strong majority also maintains that no innovation should be licensed unless science has first proved it to be safe. As if it could! The controversy over the mumps–measles–rubella (MMR) vaccine showed that large sections of the public and, worse, the media either fail to understand the importance of evidence or have no respect for it.

Second, science is one of the pillars of civilization and liberal democracy, as that eminent philosopher of science, Karl Popper, convincingly argued. It is, he said, “one of the greatest spiritual adventures man has yet known”. Because science rejects claims to truth based on authority and depends on the criticism of established ideas, it is the enemy of autocracy. Because scientific knowledge is tentative and provisional, it is

the enemy of dogma. Because it is the most effective way of learning about the physical world, it erodes superstition, ignorance and prejudice, which have been at the root of the denial of human rights throughout history, whether through racism, chauvinism or the suppression of the rights of women.

Nothing could have better illustrated the gap between cultures than literary critic F. R. Leavis’s view that science is concerned only with “productivity, material standards of living, hygienic and technological progress”.

Dick Taverne House of Lords, London SW1A 0PW, UK
e-mail: tavernd@parliament.uk

Readers are welcome to join in the debate at www.tiny.cc/opinion191

Stick as well as carrot needed to solve age-old gender bias

SIR — Your Editorial ‘The female underclass’ highlights the problems faced by women scientists in many European countries (*Nature* **459**, 299; 2009). I’d like to comment on the situation in the United States.

Taking the biological and medical sciences, for example: from 1990 to 2004, the percentage of traditional research awards from the US National Institutes of Health (NIH) allocated to women grew from a paltry 17% to just 24% (see <http://tinyurl.com/kvtvhc>). Only 19% of tenured principal investigators at the NIH are women. These figures have hardly changed over the past decade and are dishearteningly similar to those at most academic research institutions in the country (see <http://tinyurl.com/kpav3j>).

Yet there have been more female than male graduate students in these fields over the same period. In 2005 the number of doctorates awarded to women overtook the number awarded to men (see <http://tinyurl.com/>

[nemfs6](http://tinyurl.com/nemfs6)). Although women make up nearly half of all scientists nationwide, many abandon academic research after a decade.

What is happening to these female graduates, and what can explain the startling drop-off in figures? It’s simple. Report after report has documented gender bias. For example, the 2007 report from the US National Academies, *Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering*, categorically affirms bias against women applying for grants, employment and tenure. It asserts that a woman must have a significantly superior record to be rated on a par with a man. And it rejects out of hand the purported meritocracy that determines hiring, promotions and rewards in academic institutions.

The loss of women scientists has also been attributed to their relative lack of confidence in seeking positions and securing tenure (*EMBO Reports* **8**, 977–981; 2007). Of course they are less confident — a woman is only too aware of the time and energy she must invest in overcoming bias and building up a “significantly superior record”.

If we ask what has worked in those European countries that have managed to curtail destructive habits of bias and exclusion, again the answer is simple. As you point out, it takes “sticks as well as carrots”. No sensible man would give up his advantage by conceding that he is intellectually inferior to a female colleague. And no university yet seems prepared to remove men who are guilty of blatant acts of bias.

What is at stake is not only justice: it is the competitiveness of science in the United States. When half of our brightest scientists leave academic research because their intelligence and common sense tell them they are wasting their considerable skills, how can we possibly generate the best science?

Colleen E. Crangle Converspeech LLC, 60 Kirby Place, Palo Alto, California 94301, USA
e-mail: crangle@stanfordalumni.org

ESSAY

The future of saving our past

As letters and diaries give way to e-mails and laptops, fresh challenges and opportunities have emerged for archivists. **Jeremy Leighton John** explores the digital wilderness for the British Library.

In 2000, the tenaciously original evolutionary biologist and Royal Society professor Bill Hamilton died after an expedition to the rainforests of Africa, and within months his archive was delivered to the British Library in London. Unlike archives received in the past, this one did not consist only of boxes of papers — handwritten letters, typed draft essays and the like. It also included a hoard of computers and storage media from the early 1960s onwards — from 5-hole paper tapes and 80-column punched cards through to optical disks — which occupied 26 of the 200 boxes.

Bill was my mentor, and I grew to know him well, living with his family for four years while completing my PhD in evolutionary biology. At the library, his archive came to me, along with the question of how best to deal with its digital elements. Over the years, Bill's digital archive has been joined by those of other eminent researchers, including evolutionary biologist John Maynard Smith, developmental biologist Anne McLaren and computer scientist Donald Michie.

Digital revolution

The digital revolution is transforming the nature of personal archiving: from curation techniques to the kinds of lives being preserved for posterity — not just the rich and famous, but now everyone participating in the digital age. Most surviving ancient and medieval documents relate to legal, ecclesiastical and formal secular knowledge. Few bear direct witness to the details of everyday lives. Since the sixteenth and seventeenth centuries, paper has helped to open the way for widespread personal writing, printing has facilitated greater literacy and a more robust postal system has promoted distant communication. Even so, unless an archive has gone to a repository or

collector, personal papers have been kept in numbers by only the wealthiest families.

With the emergence of personal computing in the 1970s, more and more people are passing on details of their lives to future generations as digital files. It has been estimated by the International Data Corporation that by 2010

computers and devices that can make exact copies of disks — an approach that is now being adopted by archival institutions worldwide.

The original files residing on a scientist's disks and tapes will eventually become unreadable. It is therefore imperative to make exact copies of these files on fresh media in a demonstrably rigorous way. Merely turning a computer on risks altering files, so devices called write-blockers are used to prevent this from happening.

At the heart of the process is the forensic 'imaging' of an original disk. A single 'image' file, or bitstream, is made of the entire disk. It incorporates a 'map' that allows exact digital replicates of the original files to be recreated from it. Hash values are calculated for every file: short sequences of alpha-numeric characters akin to 'digital fingerprints', unique to each file. If, decades later, the same hash value is obtained for a file using the same hashing algorithm, one can be confident that the file has not been altered.

So as not to rely on software that is subject to obsolescence or is esoteric, we also create (with varying degrees of fidelity to the look and feel of the original) digital facsimiles that, through conversion to suitable file formats such as PDF or XML, are readily usable on modern computers and (let us hope) easily transferred to new computer systems.

However, the digital replicates must be preserved to retain — to the fullest extent possible — the information represented in the original files. This information will be needed should, for example, a future scholar wish to interpret precisely and accurately the original styles, layout and dynamic behaviour of a scientist's files — including home-made computer programs. These replicates can be presented with high-fidelity emulators of ancestral software and hardware.

The British Library houses a small range of classic computers with tape and disk



nearly 70% of the digital universe will be created by individuals rather than organizations. This offers an immense source of fascinating and useful personal data. But it also presents technical, ethical and legal challenges.

In 2003, I was formally appointed as the first curator of what we call eMANUSCRIPTS at the British Library, responsible for these personal digital objects and for the digital manuscripts project, which aims to research and develop the necessary procedures to curate such archives.

How can the authenticity of files be demonstrated years and centuries later? How can we ensure that a file's dates and other embedded metadata are not inadvertently changed, and are properly extracted? These kinds of considerations motivated the use of computer forensic techniques and the establishment of a secure digital scriptorium in the British Library. The facility is equipped with forensically sound

ROSEMARY WOODS

drives, which are used not only in digital capture, but also to help us understand ancestral computer systems and to appraise the fidelity of emulators.

The propensity for digital media to become unreadable, or to be protected by passwords and encryption, has led to other important changes in personal archiving. Curators are now approaching scientists directly, proactively, on their retirement or even earlier. In doing so, opportunities arise for gathering even more information.

Enhanced curation activities include audio interviews that seek a scientist's memories and life story, panoramic photography and three-dimensional graphical imagery of home and study environments, conversational video walks through the local landscape, and recordings of recollections as the person looks through photos and manuscripts. The British Library is producing interactive panoramas for the research and writing environments of eminent scientists and writers: from Fellows of the Royal Society to Poets Laureate.

Personal progress

The archived materials of influential scientists are of immense interest to historians. But life information also has the potential to be of incalculable benefit to scientific and human advancement. Diaries of travellers, letters of diplomats, logbooks of ships' officers and local family archives have long yielded geological, meteorological and sociological data. Records of measurements of magnetic north found in ship logs, for example, have been used to reconstruct the history of Earth's magnetic field.

Some important surveys have actively collected personal information: the UK National Child Development Study has over the years sought the participation of 17,000 individuals born in one week in 1958, and more recently has begun collecting DNA to look for relationships between genetic, medical and lifestyle factors. But this premeditated gathering of structured information is typically expensive or limited. In the digital era, vastly greater quantities of 'freestyle' personal information can be beneficially garnered.

A key challenge is to find a way to give scientists and bona fide researchers access without impinging on privacy. Ideally, some trusted repositories (public, non-governmental or even commercial) would ask users to what extent they would like their information, perhaps anonymized, to be available for

research. While mediating appropriate access, the repositories would help to ensure authenticity. For the most influential and creatively productive individuals, including scientists, the eMANUSCRIPTS themselves might be held for safekeeping; for others, the files — existing, as they would, in enormous quantities — might remain in the care of individuals (as 'archives in the wild', so to speak), with a repository holding only their hash values. If and when an eMANUSCRIPT becomes of interest to a scholar or scientist in the future, the hash value would vouch for its long-standing existence and integrity. Various models for future access can be imagined; the most suitable will ultimately be determined by the nature of emerging technologies and the legal and ethical contexts in which they operate.

The digital lives project — the British Library's first research project under its newly gained status as an Independent Research Organisation, funded by the Arts and Humanities Research Council — aims to further understand personal digital archives and their research potential in the twenty-first century. Through this project we are seeking to ascertain how files are generated, obtained, organized, used and stored by the public and by academics, and to understand the legal and ethical environment in which they exist.

One of the project's components is to look at the effect of the proliferation of cloud computing. People are relying increasingly on online service providers such as Google, Flickr, Facebook and Carbonite to create their files, to make them available to others or to store them remotely. Of particular relevance to the project are the web locations that are not publicly available but are restricted to the individual or friends and family — sites that typically would not be harvested by web archiving programmes. Is this information being archived at all or simply lost? If kept, who maintains legal ownership of it or could make it available for research? We are only beginning to answer these questions.

Many other important issues are arising and will do so with ever greater frequency. New forms of personal information continue to emerge. Future generations may well have comprehensive video and global-positioning-system recordings of their lives, along with records of neurological and physiological parameters such as heart rate, not to mention personal DNA sequences. Bionic devices that partially restore, enhance or extend an individual's physical or sensory capabilities will

be digitally tuned for each individual, just as digital hearing aids are today. Personal fabricators will allow individuals to create for themselves useful and ornamental physical artefacts. People already interact online through virtual versions of themselves in online gaming environments, and in future this will be advanced with immersive visualization, complemented by touch, taste and scent. What will happen to these digital representations in the long run?

Evolving archives

Perhaps the most fundamental change arising from the digital nature of personal archives lies in their passing from one generation to the next. With a paper letter or diary only one sibling could inherit the original from a parent, but today siblings are receiving identical eMANUSCRIPTs (from texts to videos).

After many generations, there will be highly diverse personal digital archives distributed throughout the population, containing many replicates of identical eMANUSCRIPTs (as well as versions that have been modified, deliberately or inadvertently). It will almost certainly be possible to create phylogenetic networks or trees from these extant personal digital archives, and even to surmise from them the composition of ancestral archives. In the digitally networked era it may be possible for the first time to capture, map and analyse in powerfully quantitative detail the movement and influences of eMANUSCRIPTs with the ideas, observations of nature and accounts of events borne by them.

It will not escape readers with an interest in evolutionary computing that such changing archives, like artificial life, evolving software and computer viruses, may provide another complex system for research — a natural history of evolving digital lives manifested in sets of eMANUSCRIPTs. Bill Hamilton's own archive makes evident his strong interest in evolutionary computing. I believe he would have been intrigued, and pleased, to find that it has been stimulating this line of thinking. ■

Jeremy Leighton John is curator of eMANUSCRIPTS at the British Library, and is principal investigator of the digital lives research project.
e-mail: jeremy.john@bl.uk

FURTHER READING

Gantz, J. F. *The Expanding Digital Universe. A Forecast of Worldwide Information Growth Through 2010* available at www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf (2007).
Garfinkel, S. L. *Int. J. Dig. Crime Forens.* **1**, 1–28 (2009).
John, J. L. *Adapting Existing Technologies for Digitally Archiving Personal Lives* available at www.bl.uk/ipres2008/programme.html (2008).
Williams, P. et al. *Digital Lives: Report of Interviews with the Creators of Personal Digital Collections* available at www.ariadne.ac.uk/issue55/williams-et-al (2008).

"Archives in the wild have the potential to be of incalculable benefit to scientific and human advancement."

BOOKS & ARTS

Looking for planets like ours

The hunt for habitable worlds near other stars brings home the realization that our own Solar System might not be as special as we think, says **Michael Brown**.

The Crowded Universe: The Search for Living Planets

by Alan Boss

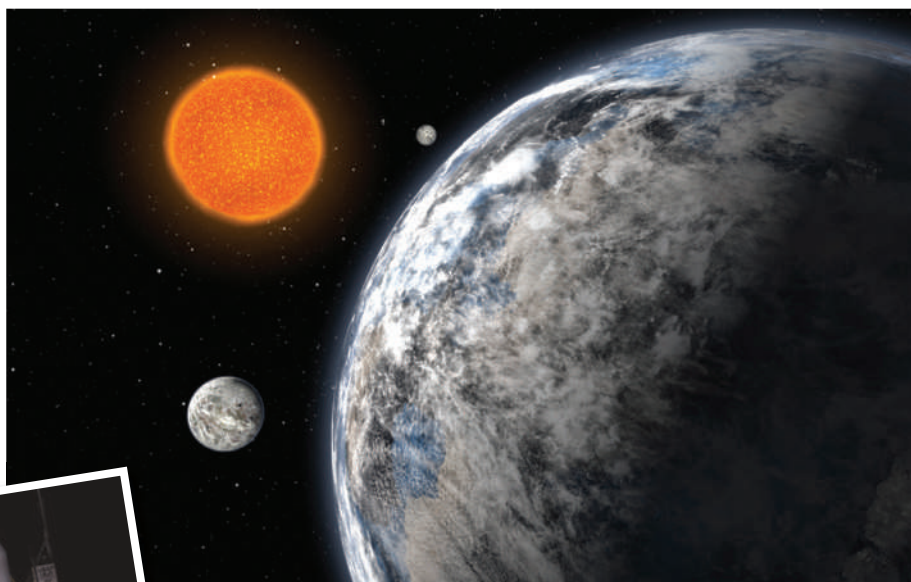
Basic Books: 2009. 256 pp. \$26, £15.99

Some 250 years ago, the philosopher Immanuel Kant laid out an account of the Universe that sounds remarkably modern. In his *Universal Natural History and Theory of the Heavens*, he declared that moons go around planets, planets go around stars and stars go around the Milky Way. The Solar System had an understandable origin, and inevitable consequences:

"The planetary structure in which the Sun at the centre makes the spheres found in its system orbit in eternal circles by means of its powerful force of attraction is entirely developed, as we have seen, from the originally distributed basic stuff of all planetary material. All the fixed stars which the eye discovers in the high recesses of the heavens and which appear to display a kind of extravagance are suns and central points of similar systems." In other words, gravity takes stuff and turns it into stars, which are surrounded by swarms of planets, and it has done so everywhere you see a star in the sky.

Until recently, this fact could be verified for only one star: the Sun. Then, in 1995, astronomers Michel Mayor and Didier Queloz announced the discovery of the first planet orbiting a star other than the Sun (*Nature* 378, 355–359; 1995). Today, almost 300 stars are known to have planets around them. It is not quite Kant's "all the fixed stars which the eye discovers", but it's getting close. It has been accepted since the seventeenth century that our Sun is not special, but is one of many stars in the Universe. Now, at the beginning of the twenty-first century, it is clear that our planets are not special either.

Except that some exoplanets are special. It is tempting to describe the many planetary systems that have been discovered so far as weird. Rather than the 'inevitable' orderly arrangement of our own Solar System — with small planets close, large planets far, and everything going around the Sun in near-circular orbits in a common disk — we have instead found that almost anything goes. Planets the size of Jupiter orbit



R. MITCHELL-RYALL, T. FARRAR/NASA/ESO

Ground-based telescopes have revealed exoplanets such as the three 'super Earths' in this artist's impression. The Kepler space telescope (left) could yield even more.



their stars at distances closer than Mercury, other planets have orbits as elliptical as some comets in the Solar System, and others lie farther from their central star than anything in our Solar System. Weird indeed.

The only type of planetary system that we have not found, it seems, is one like our own.

But the special position that our home system holds is now in jeopardy. *The Crowded Universe* tells the story of the development of NASA's Kepler space telescope, which was launched from Earth in March this year to orbit the Sun. Kepler's three-and-a-half-year mission is simple: to find the Earths. Kepler and a similar French-led mission, COROT, are the first ones with a chance to tell us whether planets like ours are as common as Kant hoped or as rare as some astronomers think.

Alan Boss, a planetary scientist at the Carnegie Institution in Washington DC, weaves the story of Kepler with the larger tale of the booming field of exoplanets. As someone whose career in astronomy has spanned the period Boss discusses, I am glad someone was taking notes. It is fun to revisit the days when each new planetary discovery was an exciting

event. Multiple teams struggled to outdo the others with firsts. First planet at the distance of the Earth! First transiting planet! First multiple planet system! It is easy to forget that most of the exoplanet field is less than a decade old.

Boss adds the insider story of the definition of planets in our own Solar System, including an account of the inner workings of the International Astronomical Union committee that decided how to classify Pluto, Eris and the other small bodies that we now call dwarf planets. The demotion of Pluto from planet class was unassailably reasonable, but the events

leading up to it were some of the more publicly comical occurrences in recent astronomical history. It is a reminder that — for all their command of the physics of the Universe — astronomers, being human, have the capacity for near-infinite folly.

But Pluto is just a distraction for Boss, and rightly so. The meat of *The Crowded Universe* is the race to find another Earth. It would have been easy for his writing to get caught up in the day-to-day minutiae of meetings and detailed descriptions of research progress. But Boss never forgets that we are privileged to live in a time when a revolution of near-Copernican magnitude is unfolding.

Even if the Kepler and COROT missions



do find an abundance of planets, the Kantian revolution will not be complete. The new planets might be exactly the same size as Earth and orbit their stars at the same distance, and although an astronomer might be willing to call such a thing Earth-like, most people will look for more. Does it have liquid water? Does it have a recognizable atmosphere? And, inevitably, could it — does it — support life?

Finding the answers to these questions will take decades. Kepler and COROT are merely steps along the way. In the meantime, we can take solace from Kant: "I am of the opinion that it is not particularly necessary to assert that all planets must be inhabited. However, at the same

time it would be absurd to deny this claim with respect to all or even to most of them."

It took nearly 250 years to prove him mostly right the first time. With a little luck and perseverance — and, as Boss shows, a lot of work by astronomers around the world — the final step may just come a little faster. ■

Michael Brown is professor of planetary astronomy in the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. e-mail: mbrown@caltech.edu

See www.nature.com/astro09 for more on the International Year of Astronomy.

Charting the heavens from China

The Dunhuang Star Chart

The British Library, London

Until 18 August 2009.

Along the ancient trade route of the Silk Road connecting China and the West, the Mogao Caves honeycomb the Mingsha Hill some 25 kilometres southeast of Dunhuang, a desert town in Gansu province. Excavated between the fourth and fourteenth centuries, the caves were Buddhist shrines and temples where travellers prayed for the success of their journeys.

In 1900, the Taoist priest Wang Yuanlu propelled the Mogao Caves to the status of an archaeological crown jewel when he stumbled upon a hidden library in Cave 17. It contained more than 40,000 manuscripts on a myriad of subjects, from religion, history, art and literature to mathematics, medicine and economics. The documents had been sealed in the cave by Buddhist monks in the eleventh century.

Among the manuscripts was an exquisite star chart. It shows the entire sky as visible from China, skilfully drawn by hand in red and black inks onto a fine, four-metre-long paper scroll. In 1907, archaeologist Marc Aurel Stein took the chart and more than 7,000 other cave manuscripts to the British Museum in London.

Dated to between 649 and 684 AD, the chart is the oldest extant graphical star atlas in the world, explains Susan Whitfield,

director of the British Library's International Dunhuang Project, which aims to make information and images about the artefacts available on the Internet. The atlas is on display at the British Library in London this summer to celebrate the International Year of Astronomy.

The atlas is divided into two sections. One shows 26 drawings of differently shaped clouds accompanied by text on cloud divination. The other section portrays 12 star maps, each depicting a 30° division of the sky in the east–west direction, plus a map of the circumpolar sky. The star positions are drawn as observed from

a latitude of 34°N, possibly from the Imperial Observatory in Chang'an (present-day Xi'an) or another site in Luoyang.

The atlas shows 1,339 stars arranged in 257 groups, or asterisms, two of which resemble the constellations of the Big Dipper and Orion. It includes faint stars that are difficult to see with the naked eye, and several in the Southern Hemisphere. The styles of the dots differentiate the three schools of astronomical tradition established during the Warring States period (476–221 BC), each of which adopted alternative names and descriptions for the star groups.

The positions of the brightest stars are surprisingly accurate to within a few degrees, says astronomer Jean-Marc Bonnet-Bidaud of the CEA, the French Atomic Energy Commission, who has studied the atlas together with Whitfield and Françoise Praderie of the Paris Observatory (J.-M. Bonnet-Bidaud, F. Praderie and S. Whitfield *J. Astron. Hist. Herit.* 12, 39–59; 2009). Stars near the celestial horizon are drawn using a cylindrical projection, in which meridians are mapped to equally spaced vertical lines, and circles of latitude are mapped to horizontal lines. The circumpolar region uses an azimuthal projection, preserving the directions of the stars from a central point. These methods are still used in geographical mapping today.

Ancient Chinese astronomers divided the celestial circle into 12 sections to follow the orbit of Jupiter, known as the Year Star in China, which loops the Sun about every 12 years. The Jupiter cycle is also the basis for the 12 months of the year that make up the Chinese calendar. On the Dunhuang chart, the text accompanying each star map names that region of sky, the astrological predictions associated with it and the states of the Chinese empire thought to be influenced by that division.

The chart may have been reproduced from an earlier atlas by tracing it on to fine paper. It has no coordinate grid, and shares wording with another traditional astronomical text, *Yue Ling*, or *Monthly Ordinances*, which has been dated to around 300 BC. Yet it remains the earliest-surviving detailed map of the entire northern sky, pre-dating others by several centuries. Older star maps described only part of the sky. *The Book of Fixed Stars*, an Arabic work written by the Persian astronomer Abd al-Rahman al-Sufi (903–986 AD), displays individual constellations but gives no information on their relative positions. The



The three stars that make up the familiar 'belt' of Orion are recognizable in this panel from the seventh-century star chart discovered near Dunhuang, China.

oldest-known star chart in Europe is the Vienna manuscript. Dated to 1440 AD, it shows only a limited number of stars in northern constellations, plotted in an azimuthal projection from the ecliptic pole.

The chart may have been used to consult the heavens to predict earthly events. Astronomy was an imperial science in ancient China, and court astronomers and astrologers created star

charts from at least the fifth century BC. Chinese emperors sought celestial clues for political and warfare decisions, and the importance of divination led to an early precision in star catalogues.

But why was the chart kept in the Mogao Caves rather than in the imperial archive? "It remains a mystery," says Whitfield. A political and secret document, it may have served a military purpose rather than being a guide for

travellers. When the Taoist priest discovered the hidden library, he could hardly have guessed that he was opening the door to a world of such fascinating antiquity. ■

Jane Qiu writes for *Nature* from Beijing.
e-mail: jane@janeqiu.com

See www.nature.com/astro09 for more on the International Year of Astronomy.

Superstition challenged

Grimoires: A History of Magic Books

by Owen Davies

Oxford University Press: 2009. 384 pp.
£14.99, \$29.95

You might expect to find grimoires — collections of magic spells, recipes and charms — on the shelves of medieval mystics or in the pages of Harry Potter books. But as social historian Owen Davies shows, they are not confined to history and fantasy.

In the 1960s, for example, the German Lutheran minister Kurt Koch waged war against what he called the "flood of magical conjuration which washes the Alps", namely the superstitions he had found across southern Germany, Austria and Switzerland. To his dismay, such beliefs were promoted in cheap, mass-produced grimoires such as *The Sixth and Seventh Books of Moses* — a title that had circulated in Europe since at least the eighteenth century.

Two contrasting positions in the modern view of grimoires are personified in another dispute in Germany in the 1950s. On one side was Johann Kruse, a schoolteacher who had seen many people, including his own mother, accused of witchcraft in rural Schleswig-Holstein in northern Germany. Like Koch, he wanted to see such beliefs banished. In 1950 Kruse founded the Archive for the Investigation of Contemporary Witchcraft Superstition, and he published the exposé *Witches Among Us?* the following year. In 1956, he successfully sued the publisher Planet-Verlag for selling a cheap version of *The Sixth and Seventh Books of Moses*. Kruse's campaign seems at first like a noble attempt to combat ignorance and deceit by targeting publishers who were exploiting the gullibility of uneducated people. But such efforts often had moralistic overtones, akin to attempts to suppress pulp fiction in favour of more 'improving' literature.

On the other side of this debate were academic folklorists such as Will-Erich Peuckert, who testified for the defence against Kruse and felt that folk beliefs were a valid part of



Johann Kruse's 1950s campaign against superstition and sorcery targeted books conveying such beliefs (inset, right).

cultural tradition. Davies is refreshingly neutral, content with wry asides that leave no doubt about his views on these childish compendiums of 'magic'.

But grimoires weren't always ridiculous. Some collections of recipes and tricks from antiquity, such as the Stockholm and Leiden papyri — discovered in the 1820s and probably made in Egypt in the third century AD — provide a valuable window on the technologies of their age, describing the preparation of medicines, pigments, dyes and metals. And some 'magic' books — such as Giambattista Della Porta's *Natural Magic* (1558), which describes a camera obscura — are scientific treatises on mathematics and optics. These claim an allegiance to magic only because of the Neoplatonic view that natural magic is the mechanical system of nature, a web of hidden or occult forces.



In this regard, Davies's book disappoints. Despite offering an overview of magical tradition, he never really beds it into the history of ideas wherein magic occupies a valid pre-scientific role. One looks in vain for the sort of synoptic theses that motivate, for example, Keith Thomas's magisterial *Religion and the Decline of Magic* (Weidenfeld and Nicholson, 1971), or Norman Cohn's study of witchcraft and persecution, *Europe's Inner Demons* (Basic Books, 1975).

Most of all, there is no sense of why magical belief proves so tenacious when magic itself does not work. As anthropologist Bronisław Malinowski argued, magic ritualizes hope in an adverse world, so that one coincidental success cancels out countless failures. And as Thomas showed, religion encouraged magical belief even while competing and sometimes merging with folk superstitions. In the Middle Ages, many uneducated parish priests conducted services as arcane rituals, with an incomprehensible liturgy and the Eucharistic host wielded like a talisman or cure-all.

Magic has long been associated with experimental science, as made clear by Lynn Thorndike in his multi-volume survey of those two activities, published between 1923 and 1958. Technology often carried the suspicion of

demonic witchcraft, not least in the invention of the printing press, which is sometimes attributed to Faust instead of Johannes Gutenberg. That is one reason why grimoires can be read as practical guidebooks, such as the 'how-to' *Kunstbuchlein* of German tradition. As William Eamon showed in *Science and the Secrets of Nature* (Princeton University Press, 1994), these made a selling point of their promise of forbidden and hidden knowledge.

Davies reveals with relish just how banal this could make such 'secret' books. They might offer nothing more dangerous than cures for stomach ache and bad breath — although few

plumb the Pythonesque depths of an Icelandic book from the seventeenth century with its runic spell to "afflict your belly with very great farting". By the twentieth century, the history of magic becomes a dispiriting, yet curiously compelling tale of charlatans and quacks who invented comically grandiose sects and titles. Davies deflates these with the understatement: "The Hermetic Brotherhood of Luxor was a mail order organization founded in 1884 by Peter Davidson, a Scottish violinmaker." He stops short of mentioning the flirtation of today's celebrities with vulgarized Kaballah — the point is made.

Also left implicit is Davies's analogy between popular grimoires and the modern plague of 'how-to' titles offering success in wealth or love. One can readily see that they are part of the same tradition, promising exclusive access to empty secrets. Internet magic has not diminished the desire for printed grimoires: given their fetishistic status, it seems that possession is still nine-tenths of the lore. *Grimoires* makes clear the continuity of magical belief in popular culture. Science and technology do not expel it, but merely give it new forms. ■

Philip Ball is a writer based in London.

The technology of illusion

Interactivos? Lima '09: Magic and Technology

Escuelab, Lima, Peru. Sponsored by Medialab Prado.

Workshop held 13 April until 13 May 2009.

Magic is not limited to the tricks performed at children's parties. It can refer to anything that resists explanation, from cognitive illusions to high-tech wizardry. This broader sense of magic was in the air in Lima, Peru, earlier this spring, when engineers and artists converged to explore the intersection of magic and technology, with awe-inspiring results.

"It's nuts, it's very ambitious," said Julian Oliver, an artist from New Zealand who helped to organize the workshop for Medialab Prado, based in Madrid, Spain. "People bring the skeletons of ideas and build them in just two weeks."

The workshop began with a digital nod to the time-honoured manipulation of playing cards and coins. Using a suitcase that folds out to become a digital overhead projector, a Spanish magician doctored a live video feed of his tabletop tricks using a pair of wireless shoes.

Other artists endowed spectators with unearthly powers. One Brazilian artist gave visitors a 'magic wand' to change the outline of their shadows. A Mexican inventor used a hidden theremin, the musical instrument that responds to the body's electrical fields, to activate a magnet that causes a pair of ghostly silver spoons to flip over when someone approached. "We had to reject a project that involved a flame you could control with your hands by way of gas injections," said Oliver.

The conception of magic in Peru was not always the art of deception seen in the West. Instead of Houdini-style illusions, "there



Viewing your arm through this box of tricks (top) transformed it into the furry equivalent shown above.

is superstition and metaphysics" in Peru involving "rites, summoning and prognostication", said Oliver. *Abracadabra Pata*, a box in which visitors saw their own arms transformed into the legs of insects or lions, gave a taste of the spirit of witchcraft. A sense of the miraculous also pulsed through an installation that asked passers-by to synchronize their heartbeats to revive an image of a dead man projected onto the floor, in a sort of shamanistic healing-ritual-turned-video-game that took its inspiration from the Peruvian poet César Vallejo.

Magic could be defined as that which science has not yet made intelligible. But "even if science could explain everything, there would still be a place for magic," noted Kiko Mayorga, a Peruvian engineer and artist who co-founded the Escuelab centre that hosted the installations. This is because "magic is linked to surprise, and one's attention can never take in everything," he wrote.

Magic can comprise inventions that baffle us until we understand them. But this ignores a crucial fact: the best illusions rely not on advances in technology, but on permanent facts of psychology. "Magic, in the traditional sense, doesn't occur in the world, it occurs inside people," explains Oliver. "The magician merely helps people to trick themselves." ■

Jascha Hoffman is a writer based in New York.

NEWS & VIEWS

PLANETARY SCIENCE

The Solar System's extended shelf life

Gregory Laughlin

Simulations show that orbital chaos can lead to collisions between Earth and the inner planets. But Einstein's tweaks to Newton's theory of gravity render these ruinous outcomes unlikely in the next few billion years.

In the midst of a seemingly endless torrent of baleful economic and environmental news, a dispatch from the field of celestial dynamics manages to sound a note of definite cheer. On page 817 of this issue, Laskar and Gastineau¹ report the outcome of a huge array of computer simulations. Their work shows that the orbits of the terrestrial planets — Mercury, Venus, Earth and Mars — have a roughly 99% chance of maintaining their current, well-ordered clockwork for the roughly 5 billion years that remain before the Sun evolves into a red giant and engulfs the inner Solar System.

The constant interplay of gravitational attractions between planets acts to degrade their repetitive and predictable motions. Over time, a system of orbits can become increasingly disordered, and, like a poorly balanced tyre that tears itself off the axle of a moving car, planets might fling each other out into space or into their parent star, or collide with each other. The census of extrasolar planets has revealed instances (such as the outer two planets orbiting the nearby star Upsilon Andromedae) where we seem to be observing a system in which one of the original planets has been ejected, leaving evidence of the catastrophe in the form of an ongoing back-and-forth exchange of angular momentum between the survivors².

To all appearances, our Solar System seems a model of stability. Phenomena such as eclipses can be pinpointed over the millennia, and the motions of the planets themselves can be

charted with confidence tens of millions of years into the future. An ironclad evaluation of the Solar System's stability, however, eluded mathematicians and astronomers for nearly three centuries.

In the seventeenth century, Isaac Newton was bothered by his inability to fully account for the observed motions of Jupiter and Saturn. The nonlinearity of the gravitational few-body problem led him to conclude³ that, "to consider simultaneously all these causes of motion and to define these motions by exact laws admitting of easy calculation exceeds, if I am not mistaken, the force of any human mind".

During the 1700s, Continental mathematicians, including Leonhard Euler, Joseph-Louis Lagrange and Pierre-Simon Laplace, developed elegant perturbative methods to describe long-term orbital evolution. Their approach met notable success with Laplace's demonstration that the so-called *grande inégalité* in the motions of Jupiter and Saturn could be attributed to the fact that the orbital period of Jupiter is close to two-fifths that of Saturn, resulting in a near 5:2 orbital resonance between them. Laplace believed that the planetary orbits would be stable and predictable for all time, an attitude that probably contributed to his formulation of a rational determinism⁴ — the belief that initial conditions and physical laws precisely determine the future.

By the 1850s, however, it was recognized that the higher-order terms in the planetary

'disturbing function' could not be neglected, and consideration of these terms revived the question of orbital stability. In 1889, Henri Poincaré demonstrated that even the gravitational three-body problem cannot be solved by analytic integration, thereby eliminating any possibility that an analytic solution for the entire future motion of the eight planets could be found. Poincaré's work anticipated the now-familiar concept of dynamical chaos and the sensitive dependence of nonlinear systems on initial conditions⁵.

In recent decades, computers have reinvigorated celestial mechanics. Orbital predictions obtained from numerical integration of the planets' equations of motion demonstrated that the planetary orbits will indeed become chaotic, with typical Lyapunov times — the time required for chaos to significantly degrade the predictability of a system — of the order of 5 million years. Statements regarding the stability of the Solar System must therefore be expressed in terms of probabilities. Computers are now fast enough to be able to produce forward models of the Solar System throughout the Sun's remaining 5-billion-year hydrogen-burning lifetime. One insight that has emerged is that, from a dynamical point of view, the Solar System is effectively two systems of planets. The gas giants — Jupiter, Saturn, Uranus and Neptune — constitute an extremely stable constellation, whereas the rocky terrestrial planets are on a far less solid footing. Were one

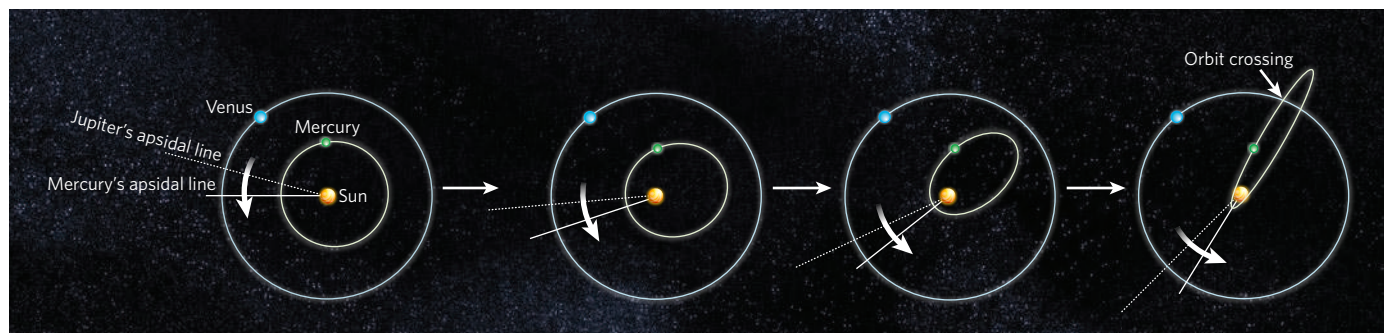


Figure 1 | Trouble with Mercury. Numerical simulations by Laskar and Gastineau¹ indicate that there is a roughly 1% chance that the inner Solar System will become destabilized during the next 5 billion years. The Solar System's Achilles heel is a secular resonance between Jupiter and Mercury, in which the orbits of the two planets, and hence their apsidal lines — the lines that connect the Sun to the point of closest approach of the planetary orbit — precess at the same rate. If the resonance is established, the eccentricity of Mercury's orbit increases steadily over a million-year timescale, and eventually crosses that of Venus. Once orbit crossing occurs, a variety of disastrous outcomes are possible, several of which are detailed by the authors¹.

to eliminate the Sun's eventual mass loss and its damaging encounters with passing stars in the far future, the outer planets would evolve with substantially unaltered orbits for about 10^{18} years before succumbing to a weak orbital resonance, in which the perturbative attractions between Jupiter, Saturn and Uranus would generate large-scale instability⁶.

Laskar and Gastineau's work¹ is the culmination of a long-running effort to establish a probabilistic assessment of the orbital stability of the terrestrial planets. They report highly detailed numerical simulations of the evolution of the whole Solar System using the most accurate available planetary ephemerides (a table of the precise positions and velocities of the planets at a specific time). The simulations indicate that Mercury, in spite of its diminutive size, poses the greatest risk to the present order. In a small but disturbing subset of possible future trajectories, Mercury becomes trapped in a 'secular resonance' with Jupiter, a state of affairs in which the elliptical figure of Mercury's orbit rotates in synchrony with Jupiter's

orbital precession (Fig. 1). If the Jupiter–Mercury resonance is established, Mercury's orbital eccentricity will increase to the point at which it intersects the orbit of Venus, setting the stage for catastrophe.

But the odds of Mercury entering a secular resonance are greatly reduced by the small modifications that Einstein's theory of general relativity imparts to the planetary motions. Famously, 43 arcseconds per century of Mercury's total precession is due to the effect of general relativity. This correction effectively detunes the Mercury–Jupiter interaction, and decreases the chance that resonance will occur in the next 5 billion years to roughly 1%. This is fortunate indeed, as the Laskar and Gastineau paper¹ also relates the precise and grisly details of one case in which Mercury's destabilization leads to a wholesale exchange of angular momentum between the inner and outer Solar System. (Readers of the paper can see for themselves the consequences of this, but suffice it to say here that Earth does not fare well in the resulting interplanetary melee.)

Laskar and Gastineau's work brings closure to one of the most illustrious and long-running problems in astronomy, and in a sense the result is as satisfying as one could wish. With 99% certainty, we can rely on the clockwork of the celestial rhythm — but with the remaining 1% we are afforded a vicarious thrill of danger. What now remains is to understand the extent to which the hand of dynamical chaos that so lightly touches our Solar System has moulded the Galactic planetary census. ■

Gregory Laughlin is in the Department of Astronomy and Astrophysics, University of California Santa Cruz, Santa Cruz, California 95064, USA.

e-mail: g laughlin@ucsc.edu

1. Laskar, J. & Gastineau, M. *Nature* **459**, 817–819 (2009).
2. Ford, E. B., Lystad, V. & Rasio, F. A. *Nature* **434**, 873–876 (2005).
3. Cohen, I. B. & Smith, G. E. *The Cambridge Companion to Newton* (Cambridge Univ. Press, 2002).
4. Laplace, P. S. *Traité de Mécanique Céleste* 5 vols (Gauthier-Villars, 1799–1825).
5. Poincaré, H. *New Methods of Celestial Mechanics* Vol. 1 (Gauthier-Villars, 1892).
6. Murray, N. & Holman, M. *Science* **283**, 1877–1881 (1999).

CELL CYCLE

Cell division brought down to size

Kenneth E. Sawin

Cells normally divide on reaching a fairly specific size, but how cell size dictates the timing of cell division remains obscure. In fission yeast, a spatial gradient of a cell-tip protein may provide an answer.

How does a cell decide that it is big enough to divide into two daughter cells? From *E. coli* to elephants, long-term sustainability depends on coordinating cell growth (an increase in total mass) with cell division. Divide too soon, and cells become progressively smaller with each generation; divide too late, and they become progressively larger. Coordination of cell size and cell division involves homeostatic mechanisms at multiple levels¹, but one question that has been particularly difficult to answer is how cell size per se affects the decision to enter mitosis — the stage in the cell cycle at which duplicated genomes separate and the cell divides in two. On pages 852 and 857 of this issue, two studies, one by Martin and Berthelot-Grosjean² and the other by Moseley *et al.*³, independently provide evidence that in fission yeast (*Schizosaccharomyces pombe*) a spatial gradient of a protein emanating from cell tips may regulate mitotic entry, thereby directly linking cell size with cell division.

Pioneering work⁴ on the cell cycle in the 1980s revealed that the key enzyme regulating entry into mitosis is cyclin-dependent kinase (CDK), a complex made up of the protein kinase Cdk1 (also known as Cdc2) and its regulatory subunit, cyclin. Protein kinases phosphorylate other proteins, thereby changing

their properties. Active CDK, with the help of several other protein kinases, phosphorylates multiple substrates to drive the events of mitosis and cell division. During interphase — the stage of the cell cycle preceding mitosis — CDK activity is held in check by several mechanisms⁵, including inhibitory phosphorylation of Cdk1 by another protein kinase, Wee1. Entry into mitosis accordingly depends on a protein phosphatase, Cdc25, that removes the inhibitory phosphorylation from Cdk1. Wee1 is itself inhibited by two related protein kinases, Cdr1 (also known as Nim1) and Cdr2 (refs 6, 7).

The unique properties of fission yeast make it an ideal model for cell-cycle research. Fission-yeast cells are cylindrical and maintain a constant width as their tips grow; as a result, cell length reflects both cell-cycle stage and cell size. Increasing the amount of the Cdk1 inhibitor Wee1 causes fission-yeast cells to divide at increased lengths, whereas increasing the amount of the Cdk1 activator Cdc25 leads to cell division at decreased lengths⁴. That these cell-cycle regulators can change the length 'set-point' for entry into mitosis suggests that cell size might be linked to mitotic entry by modulating the balance of Wee1 and Cdc25 activities in growing cells.

But how might this modulation occur?

Martin and Berthelot-Grosjean² and Moseley *et al.*³ harness the power of yeast genetic analysis to pinpoint the involvement of another protein kinase, Pom1 (ref. 8). Both groups^{2,3} show that when the *pom1* gene is deleted, cells divide at a slightly reduced length. Increasing the amounts of Pom1 increases the length at which cells divide, implying that Pom1 inhibits entry into mitosis. Additional experiments indicate that Pom1 exerts its effects on division length by affecting the Cdr2–Cdr1–Wee1 pathway, possibly through Pom1 phosphorylating Cdr2, which Martin and Berthelot-Grosjean² show occurs *in vitro*. Inhibitory phosphorylation of Cdr2 would prevent inhibition of Wee1, allowing active Wee1 to inhibit Cdk2 and delay cell division.

The link between Pom1 and Cdr2 becomes even more intriguing on observing the cellular localization of these proteins. Cdr2 is found in unusual structures known as interphase nodes, which form a band in the cortex (the region just under the cell membrane) in the middle of the cell, overlying the cell nucleus⁹. Until now, these nodes had been thought to function primarily in linking the position of the nucleus to the cell-division plane during cytokinesis — the physical separation of daughter cells after mitosis¹⁰. However, the new work^{2,3}, particularly that by Moseley *et al.*³, shows that Cdr1 and Wee1, as well as several other proteins, are also localized to these nodes. Moreover, when the *cdr2* gene is deleted, most or all of the proteins are no longer found in nodes. Thus, the medial cell cortex contains a Cdr2–Cdr1–Wee1 regulatory network, whose spatial positioning is determined by Cdr2.

During interphase, Pom1 is localized to cell tips, in the cell cortex⁸. Martin and Berthelot-

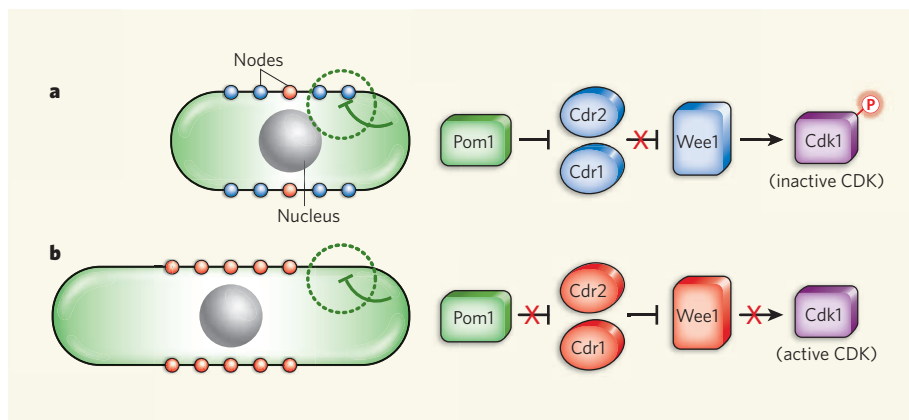


Figure 1 | A spatial gradient links cell size to cell division in fission yeast^{2,3}. The Pom1 protein kinase (green) is localized to the cell cortex, with the highest concentration at the cell tips. The cell-cycle regulators Cdr2, Cdr1 and Wee1 are present in cortical nodes in the middle of the cell (blue and red dots). **a**, In small cells, the Pom1 gradient reaches most of the cortical nodes (blue dots). Pom1 inhibits Cdr2, preventing Cdr2 and Cdr1 from inhibiting Wee1, and allowing Wee1 to phosphorylate Cdk1, thus inactivating cyclin-dependent kinase (CDK) activity and preventing entry into mitosis. **b**, In long cells, the Pom1 gradient does not reach the cortical nodes (red dots), and therefore Cdr2 and Cdr1 remain active in the nodes. Cdr2 and Cdr1 inhibit Wee1, preventing phosphorylation of Cdk1 and thereby leading to activation of CDK and mitotic entry. (This simplified diagram omits several other regulators of CDK activity.)

Grosjean² and Moseley *et al.*³ show that Pom1 distribution strongly influences the localization of Cdr2 and the other cortical node proteins. Targeting of Pom1 to the middle of the cell reduces the amount of Cdr2 at the medial cortical nodes and is accompanied by an increase in division length. By contrast, when *pom1* is deleted, the nodes are no longer restricted to the middle of the cell but rather distribute throughout the cortex of one half of the cell. Together with Cdr2 phosphorylation, this spatial regulation may contribute to control of Cdr2 by Pom1, although additional mechanisms must also regulate node distribution, because in the *pom1*-deleted cells active cell growth prevents the nodes from spreading to the other half of the cell.

These and other experiments lead Martin and Berthelot-Grosjean and Moseley *et al.* to propose a model for size control of cell division in which tip-localized Pom1 generates a gradient of cortical Pom1 that regulates Cdr2 activity (Fig. 1). The idea is strikingly simple: in small cells the 'tail' of the Pom1 gradient overlaps with the medial Cdr2–Cdr1–Wee1 network and there is sufficient Pom1 to inhibit Cdr2, allowing active Wee1 to prevent mitotic entry (Fig. 1a). However, as the cell grows and the distance between the tips and medial region increases, a crucial length is reached at which the Pom1 gradient no longer overlaps sufficiently with Cdr2–Cdr1–Wee1; Pom1 can therefore no longer inhibit Cdr2, which is free to inhibit Wee1, leading to activation of CDK and entry into mitosis (Fig. 1b).

The model is elegant, but there remain several unanswered questions. How exactly does Pom1 inhibit Cdr2? Does it phosphorylate Cdr2 *in vivo*, and, if so, what are the consequences of this phosphorylation? In principle, Pom1 could inhibit Cdr2 not only by changing

its cellular localization or cortical dynamics, but also by directly affecting Cdr2's protein-kinase activity, or by altering the interaction of Cdr2 with other regulators.

The interphase cortical nodes are also of interest — are they 'master integrators' of mitosis with cytokinesis? Cells deleted for *pom1* divide asymmetrically⁸, owing to the abnormal distribution of the nodes and their

role in cytokinesis. Because *pom1* deletion mislocalizes Cdr2, and Cdr2 is necessary for the localization of many proteins to the interphase nodes, I would predict that deletion of *cdr2* might restore symmetrical division to *pom1*-deleted cells. More experiments of this nature may give further insight into the regulation of cytokinesis.

Finally, it should not be ignored that, even in fission yeast, where cell geometry lends itself to the proposed Pom1–Cdr2–Cdr1–Wee1 mechanism, this is unlikely to be the only way of linking cell size to cell division¹. Indeed, previous work on size control in fission yeast has implicated the 'Cdc25 branch' of Cdk1 regulation in this process¹¹, so it may be time to revisit this question as well.

Kenneth E. Sawin is at the Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3JR, UK.
e-mail: ken.sawin@ed.ac.uk

1. Jorgensen, P. & Tyers, M. *Curr. Biol.* **14**, R1014–R1027 (2004).
2. Martin, S. G. & Berthelot-Grosjean, M. *Nature* **459**, 852–856 (2009).
3. Moseley, J. B., Mayeux, A., Paoletti, A. & Nurse, P. *Nature* **459**, 857–860 (2009).
4. Nurse, P. *Nature* **344**, 503–508 (1990).
5. Morgan, D. O. *The Cell Cycle: Principles of Control* (New Science Press, 2007).
6. Breeding, C. S. *et al.* *Mol. Biol. Cell* **9**, 3399–3415 (1998).
7. Kanoh, J. & Russell, P. *Mol. Biol. Cell* **9**, 3321–3334 (1998).
8. Bähler, J. & Pringle, J. R. *Genes Dev.* **12**, 1356–1370 (1998).
9. Morrell, J. L., Nichols, C. B. & Gould, K. L. *J. Cell Sci.* **117**, 5293–5302 (2004).
10. Vavylonis, D., Wu, J.-Q., Hao, S., O'Shaughnessy, B. & Pollard, T. D. *Science* **319**, 97–100 (2008).
11. Rupeš, I. *Trends Genet.* **18**, 479–485 (2002).

ECOLOGY

Traits of plant invaders

Tim Seastedt

Many plants introduced to new habitats have fewer microbial pathogens than when in their home range, and have the ability to grow rapidly. Such a combination may make for especially troublesome immigrants.

The study of invasive species is a lively, not to say contentious, area of basic and applied ecology. It remains difficult to identify the mechanisms that allow plants and animals to become abundant in new landscapes. And debate continues over whether drawing a distinction between 'native' and 'non-native' species helps in analysing the factors that affect species change^{1,2}. In a paper in *Proceedings of the National Academy of Sciences*, Blumenthal and colleagues³ present an assessment of invasive-plant traits that bears on both of these issues.

Most plant species identified as problematic and invasive are reported to have substantially fewer viral and fungal pathogens in introduced habitats compared with the pathogen load in their native habitats⁴. This alone can confer a

competitive advantage in new environments. Blumenthal *et al.* looked at 243 species of European plant that have become established in the United States. They show that the species that have probably lost the most viruses and fungi are also likely to be species capable of rapid growth, and are often good competitors under conditions of high resource availability. The authors argue that this combination of release from enemies and inherent growth ability produces a synergistic effect that explains patterns seen in the abundance and types of invaders, and in the patterns of invasion seen in many ecosystems.

Ecologists often group plants by the ability of species to persist in either resource-rich environments ('competitors') or resource-poor environments ('stress tolerators'). A second

comparison separates species able to exploit physical disturbances ('ruderals') from the other two groups⁵. When Blumenthal *et al.*³ compared pathogen numbers on these plants, an unexpected finding was that, in their home range, plants classified as competitors averaged four times as many pathogenic fungi and about seven times as many viruses as plants adapted as stress-tolerators or as stress-tolerant ruderals.

The authors then looked at the question of what best explains the number of pathogen species as well as the subsequent loss of this pathogen burden when plants move to new environments. They used a scoring procedure of plant characteristics that ranked species as competitors, stress tolerators, ruderals or intermediates. They also ranked resources (light, water, nitrogen) from sample locations. Plant characteristics best explained numbers of pathogens as well as the losses in numbers of pathogens; the pathogen burden, and its subsequent reduction, was related to resource availability.

Plants adapted to moderate- and high-resource communities are able to grow rapidly when presented with resource opportunities⁶. But the fact that such plants exhibit these growth characteristics in their home environment while possessing relatively high numbers of viral and fungal pathogens is surprising. Clearly, the plants can keep pathogens in check while competing in relatively high-resource areas. In contrast, stress-tolerant plants that persist in environments that are 'too dry' or 'too wet' do so while hosting fewer pathogens. The causal relationships among these factors can be only speculative. But it seems that a resource-rich environment allows a plant to carry a higher pathogen load and persist, whereas in low-resource areas natural selection favours only those that maintain reduced pathogen loads. Plants with a growth strategy that provides quick benefits from their tissues can tolerate a higher pathogen load; in low-resource areas they need tissues that will discourage pathogen presence.

These interpretations seem to be consistent with the principles of leaf and plant economics^{6,7} that govern the adaptive balance between growth and defence. Plant growth rates are strongly correlated with higher specific leaf area (thinner leaves) and higher foliar nutrients, and invasive plants are more likely to have these traits and therefore to grow faster⁸. Thus, instead of two separate mechanisms — high growth favours invaders; escape from enemies favours invaders — Blumenthal *et al.*³ find that both of these traits are embodied in plants that become successful invaders.

This study did not rank invaders in terms of their ability to achieve extreme dominance or 'monoculture' in their new environments, and it would be interesting to know which plants — the competitors or the ruderals — have become the 'worst' invaders. Other points to consider are that viruses and fungi are not on the radar screens of many ecologists, and for

them, 'escape from enemies' evokes images of escape from vertebrate and invertebrate herbivores. The 'enemies list' for these plants contains many not included in Blumenthal and colleagues' analysis; in escaping from microbial pathogens, many species will have escaped from their specialist herbivores as well⁹. Finally, the authors' model does not consider the 'new friends list' — the beneficial, free-living microbes or symbionts that are acquired in the new environment, and that can also provide positive feedback to plant growth¹⁰.

Invasive plants are not unique in terms of their growth physiologies; rather, these species often group within the set of plants associated with rapid-growth strategies⁸. Using this criterion, the argument that invasive species are fundamentally different from native species is not justified. But Blumenthal and colleagues' analysis³ shows that non-native species combine two characteristics — rapid growth and a lower pathogen burden (and, presumably, fewer herbivores) — that distinguishes them from native species. Thus, the combination of the two traits does set this group apart, with a subset of them possessing

extreme growth and competitive potential.

Accordingly, these results constitute good evidence for the existence of a subset of non-native species that are unique in their invasive abilities. They justify the use of the 'non-native' label in discussions on environmental management and biodiversity concerns, and more generally are a further step in the maturation of invasion biology as a predictive science. ■

Tim Seastedt is in the Department of Ecology and Evolutionary Biology, and at INSTAAR, University of Colorado, Boulder, Colorado 80309, USA.
e-mail: timothy.seastedt@colorado.edu

1. Warren, C. R. *Prog. Hum. Geogr.* **31**, 427–446 (2007).
2. Richardson, D. M., Pyšek, P., Simberloff, D., Rejmánek, M. & Mader, A. D. *Prog. Hum. Geogr.* **32**, 295–298 (2008).
3. Blumenthal, D., Mitchell, C. E., Pyšek, P. & Jarošík, V. *Proc. Natl Acad. Sci. USA* **106**, 7899–7904 (2009).
4. Mitchell, C. E. & Power, A. G. *Nature* **421**, 625–627 (2003).
5. Grimm, J. P. *Am. Nat.* **111**, 1169–1194 (1977).
6. Westoby, M., Falster, D. S., Moles, A. T., Vesk, P. A. & Wright, I. J. *Annu. Rev. Ecol. Syst.* **33**, 125–159 (2002).
7. Wright, I. J. *et al. Nature* **428**, 821–827 (2004).
8. Leishman, M. R., Haslehurst, T., Ares, A. & Baruch, Z. *New Phytol.* **176**, 635–643 (2007).
9. Keane, R. M. & Crawley, M. J. *Trends Ecol. Evol.* **17**, 164–170 (2002).
10. Mitchell, C. E. *et al. Ecol. Lett.* **9**, 726–740 (2006).

IMMUNOLOGY

Immunity's ancient arms

Gary W. Litman and John P. Cannon

Diverse receptors on two types of cell mediate adaptive immunity in jawed vertebrates. In the lamprey, a jawless vertebrate, immunity is likewise compartmentalized but the molecular mechanics are very different.

About 500 million years ago, vertebrates divided into two forms marked by the possession or lack of jaws. Today, the jawed forms are by far the more common and include organisms as diverse as humans and sharks. Immunology research has understandably centred on the jawed vertebrates, in which it has long been known that adaptive (inducible) immunity is mainly accounted for by two distinct lineages of lymphocyte: T cells and B cells. But any understanding of vertebrate immunity has to consider our distant jawless cousins, which today are represented by only the lamprey and hagfish. On page 796 of this issue, Gou and colleagues¹ do just that, and extend our concepts of the two-component nature of immune function² through their investigation of the cellular basis of immunity in the lamprey.

The business ends of T and B cells are the components that recognize foreign antigen — respectively, the T-cell antigen-binding receptor (TCR), and the B-cell receptor, which is a membrane-anchored immunoglobulin that is subsequently released as a circulating antibody (Fig. 1a). Extraordinary diversity can be generated in both TCRs and immunoglobulins outside the germline through somatic

recombination, and also through hypervariation, in the case of immunoglobulins. Consequently, T and B cells can produce an immune response to all manner of invaders. One major (cellular) type of T-cell response involves killing host cells infected by viruses or bacteria. By contrast, the (humoral) B-cell-derived response consists of the production of antibodies that largely recognize and ultimately help to destroy aliens in the bloodstream.

But jawless vertebrates have much to offer immunologists. Lampreys have been shown to mount an adaptive form of specific humoral immunity to Gram-negative bacteria, as well as to a determinant unique to human type O red blood cells³. Despite the antigenic complexity of the red-blood-cell surface, induced molecules that agglutinated type O blood cells did not agglutinate type A or B cells. Furthermore, the molecular size of the reactive molecules was inconsistent with that of the immunoglobulins³. Experimental strategies, based on knowledge of the genes encoding immunoglobulins and TCRs, failed to detect the evolutionary equivalents (orthologues) of these molecules in jawless vertebrates. These negative results became understandable with the recognition that

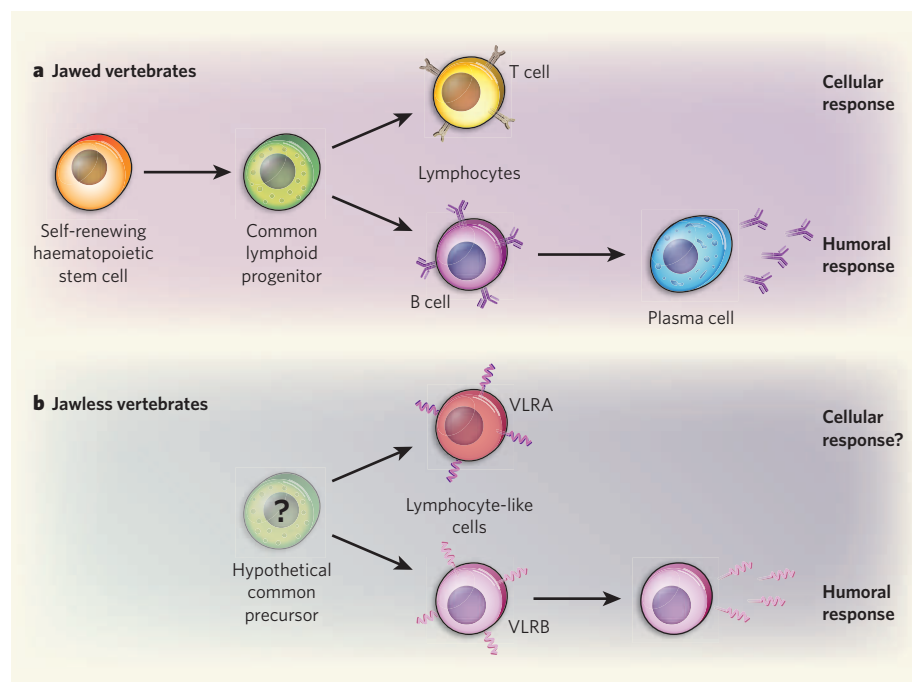


Figure 1 | Compartmentalization of adaptive immunity in vertebrates. **a**, In jawed vertebrates, a haematopoietic stem cell differentiates into a common lymphoid progenitor, and then into three lineages: T cells, which are responsible for the cellular immune response; B cells, responsible for the humoral response; and natural killer cells (not shown). On T cells, the receptor is the T-cell antigen-binding receptor and on B cells it is membrane-bound immunoglobulin. Alternative RNA processing changes membrane-bound immunoglobulin to a secreted antibody of the same specificity in a plasma cell. **b**, In jawless vertebrates, an as-yet-undefined precursor is thought to give rise to two lineages of lymphocyte-like cells defined by their expression of variable lymphocyte receptor (VLR) A or B. As Guo *et al.*¹ show in the lamprey, VLRA is expressed only on the cell surface; VLRB is expressed initially on the cell surface and is also secreted. The molecular mechanisms by which adaptive immunity is compartmentalized into cellular and humoral responses in jawed and jawless vertebrates are fundamentally different. But in both cases there was evidently evolutionary pressure to create specific lineage-restricted immune systems.

adaptive immune specificity in the lamprey and hagfish is mediated by variable lymphocyte receptors (VLRs) that are unrelated to immunoglobulin-type molecules⁴.

Much is now known about the structure and genetic diversification of VLRs, which consist of variable and invariant portions. On the basis of this, they are classified as VLRA and VLRB^{4,5}, and the complex immune repertoires of both types are somatically derived through a gene-conversion-like mechanism that recombines relatively short DNA sequences encoding leucine-rich repeats^{6,7}. A very different form of somatic recombination gives rise to diversified immunoglobulins and TCRs in the lymphocytes of jawed vertebrates.

But VLRs are produced by cells in the lamprey that morphologically resemble the lymphocytes seen in jawed vertebrates^{8,9}. Gou *et al.*¹ illustrate further parallels between VLR-bearing cells and conventional T and B cells. First, transfection of complementary DNAs encoding VLRA and VLRB into mammalian cell lines revealed that VLRB is both expressed at the cell surface and secreted, whereas VLRA is only surface expressed (Fig. 1b). The tissue distribution of VLRA is more anatomically restricted than that of VLRB. A known stimulant of T-cell proliferation in jawed

vertebrates acts on VLRA-expressing cells to a greater degree than on VLRB-expressing cells. Putative orthologues of transcription products expressed in conventional T and B cells are associated with cell populations that are respectively restricted to VLRA and VLRB expression. Finally, two types of AID/APOBEC deaminase enzyme, both of which could potentially be involved in the generation of somatic variability⁷, are expressed selectively in the two lineages.

However, the relationship of the VLRA- and VLRB-expressing cell populations to the lymphocyte lineages in jawed vertebrates is far from clear. Confounding factors are the absence of conventional T- and B-cell receptors, which are the defining lineage-specific markers, and of their respective co-receptors. And there are no equivalents of those members of the major histocompatibility complex (MHC) family that interact specifically with T cells in jawed vertebrates.

The results of Gou *et al.*¹ constitute a convincing case for the compartmental differentiation of immune functions in jawless as well as in jawed vertebrates. But the possible relationships between the VLR lymphocyte-like cells and the mediators of adaptive immunity in jawed vertebrates have probably



50 YEARS AGO

A general investigation has been in progress for some time in this Department into certain aspects of the chemistry of *Ulex europaeus*. This common furze has a local reputation as a supplementary animal foodstuff ... No previous results are available on the carotene content of the green matter (spines) of furze and its seasonal variation, which are now reported ... From October to February or March — when furze is fed to stock — the average carotene content is 126 mgm./kgm., indicating that it is a very good, and freely available, source of carotene. The relatively high values during the summer months point to the possibility of using dried furze meal as a supplementary source of carotene in animal feeding. The problem of retention of carotene in dried furze is being investigated.

From *Nature* 13 June 1959.

100 YEARS AGO

The interesting problem of the dew-pond still awaits a definite solution. That these ponds are mainly fed by mist, and not dew, can hardly be doubted by anyone who has visited them at night, situated as they are on the topmost ridges of the Downs. In the driest summer the prevailing south-west wind, as it comes up from the sea, forms on these heights after dark thick clouds of mist, which soak everything that comes into contact with them ... The source of the water in these ponds, therefore, seems evident, but the mechanism by which the mist is precipitated into the ponds is not so apparent ... It appears to me that the only possible explanation is that the particles of mist must bear charges of electricity differing in potential from that of the earth. The charge on the earth would, of course, be most dense at the summits of the hills. Hence the tendency for the mist to deposit on the top of the ridge.

From *Nature* 10 June 1909.

50 & 100 YEARS AGO

been obscured by the passage of evolutionary time. Given the extent of the evolutionary separation between jawed and jawless vertebrates, and what we know of the extraordinary levels of variation in genomic organization and the unique mechanisms associated with the rearranging antigen-binding receptors in the jawed vertebrates¹⁰, the use of a different molecule to carry out similar functions is not surprising. Whether or not the VLR system can create repertoires as complex as those seen with antibodies and TCRs remains an open question.

Ultimately, VLR diversity may well not match the extraordinarily fine specificity effected through immunoglobulins. But allelic exclusion, which restricts the rearrangement of TCRs and immunoglobulins to a single chromosome in jawed vertebrates, thus ensuring a single specific immune receptor on a lymphocyte, has also been demonstrated in lamprey VLRs^{4,6}. Other defining characteristics of jawed vertebrate lymphocytes, including clonal selection of somatic cells, lymphocyte memory, affinity maturation and intercompartmental cooperativity, are potentially common to adaptive immunity in all vertebrates.

Overall, the new findings¹ cause us to rethink the nature of immune function outside what now seems to be the ever-more-specialized case of T- and B-cell immunity in jawed vertebrates. The emerging theme reinforced here is that there was a compelling need for the evolution of effective host mechanisms to combat the elaborate immune-evasion strategies of pathogens. Adaptability in response to these strategies is achieved through integration of basic physiological processes, such as selective mobilization of DNA-repair mechanisms and differential processing of RNA¹⁰, with surprisingly few unique innovations. If the differences in the molecular basis of immune-receptor function between a lamprey and a human (or jawed fish) are as great as they seem, then there is no predicting what other adaptations may have arisen in the animal and plant kingdoms. Studies of immunity in different organisms will continue to provide examples of some of the most complex patterns of system integration and specialization in biology. ■

Gary W. Litman and John P. Cannon are at the Children's Research Institute, St Petersburg, Florida 33701, USA.
e-mail: litmang@allkids.org

1. Gou, P. *et al.* *Nature* **459**, 796–801 (2009).
2. Cooper, M. D., Peterson, R. D. A. & Good, R. A. *Nature* **205**, 143–146 (1965).
3. Pollara, B., Litman, G. W., Finstad, J., Howell, J. & Good, R. A. *J. Immunol.* **105**, 738–745 (1970).
4. Pancer, Z. *et al.* *Nature* **430**, 174–180 (2004).
5. Alder, M. N. *et al.* *Science* **310**, 1970–1973 (2005).
6. Nagawa, F. *et al.* *Nature Immunol.* **8**, 206–213 (2006).
7. Rogozin, I. B. *et al.* *Nature Immunol.* **8**, 647–656 (2007).
8. Finstad, J. & Good, R. A. *J. Exp. Med.* **120**, 1151–1168 (1964).
9. Mayer, W. E. *et al.* *Proc. Natl Acad. Sci. USA* **99**, 14350–14355 (2002).
10. Litman, G. W., Cannon, J. P. & Dishaw, L. J. *Nature Rev. Immunol.* **5**, 866–879 (2005).

ORGANIC CHEMISTRY

Synthetic lessons from nature

Huw M. L. Davies

Most syntheses in organic chemistry target one molecule, or a few closely related analogues at most. But by taking a leaf from nature's book, the latest synthetic strategy opens the door to a whole family of compounds.

Naturally occurring compounds — known as 'natural products' to chemists — have had a long-standing role in the process of drug discovery. Many of them have useful biological activities and have become the foundations of new classes of therapeutic agents. A severe constraint on such efforts, however, is the limited availability of many natural products, which are often present in vanishingly small quantities in their native organisms. The *de novo* synthesis of natural products is therefore a crucial area of research, because it can increase the availability of a compound and enable the preparation of various analogues that might have better activity than the naturally occurring molecule. On page 824 of this issue, Chen and Baran¹ describe a tremendous advance for the field: a biomimetically inspired strategy for the rapid construction of a family of 'terpene' natural products, starting from one common intermediate.

Natural-product synthesis is a well-established discipline, spanning more than a century of intense investigation². But although a wide variety of structurally complex natural

products have now been synthesized, most routes typically target a specific compound (rather than families of compounds) and involve an impracticably large number of steps. A major focus of present-day organic synthesis is therefore the development of highly streamlined and practical procedures for the rapid construction of diverse, complex natural products.

Baran's group has previously concentrated on designing such strategies for the rapid construction of alkaloid (nitrogen-containing) natural products^{3,4}, but terpenes — compounds that have hydrocarbon skeletons decorated with oxygen-containing groups — offer different synthetic challenges. Terpenes constitute the largest class of natural products, finding roles as scents, flavours and pharmaceuticals. Chen and Baran¹ recognized that, compared with typical syntheses in the lab, the biosyntheses of terpenes are highly streamlined. Relatively simple hydrocarbon cores, free of other chemical groups, are first generated using a standard set of cyclase enzymes. Structural diversity is then introduced through a series of oxidation

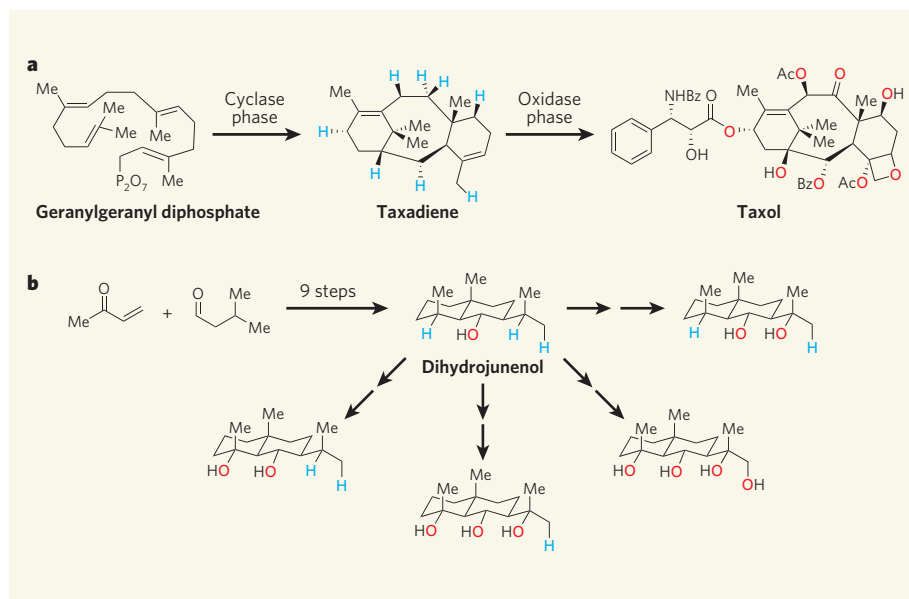


Figure 1 | Biological and biomimetic syntheses of terpenes. **a**, The biosyntheses of terpene molecules are two-phase processes, as shown here for Taxol, a naturally occurring anticancer agent. In the first phase, a precursor molecule (geranylgeranyl diphosphate) is converted by cyclases into a hydrocarbon skeleton (taxadiene). Oxidases then replace selected hydrogen atoms (blue) in the skeleton with oxygen-containing groups (red) to form Taxol. Me is a methyl group; Bz is a benzoyl group; Ac is an acetyl group. **b**, Chen and Baran¹ mimicked this approach in their chemical syntheses of eudesmane terpenes. In the first phase, they made dihydrojunenol, a simple terpene that forms the skeleton of several other compounds. They then converted one or more hydrogen atoms of dihydrojunenol into hydroxyl groups in a series of site-specific oxidation reactions, to make four more natural products.

reactions, which add oxygen-containing groups at specific sites on the hydrocarbon core.

The authors¹ highlight the differences between biological and laboratory syntheses of terpenes by considering Taxol (paclitaxel) — a terpene used as an anticancer agent (Fig. 1a). Taxol is a structurally complex molecule, and laboratory syntheses of it are lengthy and inefficient, generating only small (milligram) quantities of material. Yet the compound can be isolated from its natural source (Pacific yew trees) in greater quantities; indeed, an analogue of Taxol, which contains most of the structural complexity of the anticancer agent, is isolated annually in tonne quantities from European yew, and is used in the mass production of the drug. If a similar strategy to terpene biosynthesis could be used for the laboratory preparation of Taxol and other terpenes, not only would the resulting chemical syntheses be short and economically practical, but the inherent flexibility of the approach would offer opportunities for the preparation of a diverse range of terpene natural products.

Organic chemists typically make compounds by controlling the reactivity of 'functional groups' — reactive sites within the molecule. These usually contain unsaturated, polar or weak bonds. But the bulk of an organic molecule consists of strong, non-polarized carbon-carbon and carbon-hydrogen (C-H) bonds, which are stable to most reaction conditions. Many current research programmes seek to challenge established synthetic dogma by developing reactions involving C-H bonds⁵⁻⁸. Because these bonds are ubiquitous in organic molecules, the ability to target reactions to specific C-H bonds is a formidable challenge for organic chemists. Chen and Baran's approach¹ to making terpenes is especially audacious because it requires several sequential C-H oxidation reactions. Each of these must be deployed with exquisite control, as the molecules involved contain several sites that are susceptible to such reactions. Oxidations at the wrong sites would lead to the formation of side products and reduce the overall yield of the product.

Chen and Baran illustrate the principle of their synthetic strategy by making five members of the eudesmane class of terpene (Fig. 1b). They begin by rapidly preparing the natural product dihydrojunenol, in a practical number of steps and overall yield, thus providing plenty of material from which the remaining targets could be made. Dihydrojunenol contains only one functional group (a hydroxyl group, OH), and so its preparation mimics the cyclase phase of terpene biosynthesis. The authors then use the hydroxyl group as a chemical handle to control various site-selective C-H oxidations, thus mimicking the second phase of terpene biosynthesis. In this way the authors generate, in very few synthetic steps compared with traditional synthetic approaches, four more eudesmane natural products.

The spectacular efficiency of Chen and

Baran's synthetic strategy for making the eudesmane terpenes underscores the potential of site-selective C-H functionalization reactions for organic synthesis. The authors' approach relies extensively on the influence of the dihydrojunenol hydroxyl group to direct oxidation reactions to particular sites in the molecule, but new methods are emerging for catalytic C-H functionalization reactions that do not require directing groups on the substrates^{5,9}. As these methods mature, they are likely to become pivotal to commonly used strategies for streamlining the total synthesis of natural products and their derivatives. ■

Huw M. L. Davies is in the Department of Chemistry, Emory University, 440 Atwood Hall,

Atlanta, Georgia 30322, USA.

e-mail: hmdavie@emory.edu

1. Chen, K. & Baran, P. S. *Nature* **459**, 824-828 (2009).
2. Nicolau, K. C. & Sorensen, E. J. *Classics in Total Synthesis* (Wiley VCH, 1996).
3. Richter, J. M. *et al.* *J. Am. Chem. Soc.* **130**, 17938-17954 (2008).
4. Shen, R. A., O'Malley, D. P. & Baran, P. S. *Acc. Chem. Res.* **42**, 530-541 (2009).
5. Chen, M. S. & White, M. C. *Science* **318**, 783-787 (2007).
6. Fleming, J. J., McReynolds, M. D. & Du Bois, J. J. *Am. Chem. Soc.* **129**, 9964-9975 (2007).
7. Davies, H. M. L., Dal, X. & Long, M. S. *J. Am. Chem. Soc.* **128**, 2485-2490 (2006).
8. Wender, P. A., Hilinski, M. K. & Mayweg, A. V. W. *Org. Lett.* **7**, 79-82 (2005).
9. Davies, H. M. L. & Manning, J. R. *Nature* **451**, 417-424 (2008).

CELL BIOLOGY

The proteasome assembly line

Kiran Madura

The assembly of the proteasome — the cellular machine that eliminates unwanted proteins — is a carefully choreographed affair, involving a complex sequence of steps overseen by dedicated protein chaperones.

The proteasome degrades faulty or superfluous intracellular proteins, which are targeted to it for destruction by the attachment of the small ubiquitin protein. This multi-protein complex consists of a barrel-shaped core particle, which enzymatically degrades proteins, bound to one or two outer regulatory particles. The regulatory particle binds and unfolds ubiquitin-tagged proteins and transfers them to the catalytic chamber of the core particle. Whereas the assembly pathway of the core particle is quite well understood, little is known about how the regulatory particle is put together. That situation changes with the publication of two exciting papers in this issue, one by Roelofs *et al.*¹ (page 861) and an accompanying paper by Park *et al.*² (page 866), which report the details of regulatory-particle assembly in the yeast *Saccharomyces cerevisiae*.

The core particle of the proteasome consists of four heptameric rings: two inner rings of β -type subunits and two outer rings of α -type subunits. Although the structure of the regulatory particle has not been described in detail, it is known to consist of two parts — the base and the lid. The base is made up of six ATPase subunits (Rpt1-6) and three non-ATPase subunits (Rpn1, Rpn2 and Rpn13). The six ATPase subunits assemble into a ring, with the carboxy-terminal tails of the component Rpt proteins anchored in pockets in the α -subunits of the core particle^{3,4}.

Until recently, it was believed that proteasome assembly was a relatively straightforward process, in which an intact core particle binds to two intact regulatory particles to generate the dumbbell-shaped proteasome. We now

know that the assembly of the core particle involves multiple steps and protein-assembly factors (chaperones), which promote the correct arrangement of core-particle subunits and prevent inactive endpoints⁵.

Roelofs *et al.*¹ and Park *et al.*² reveal that, in the yeast proteasome, regulatory-particle assembly is also a complex process. They suggest that assembly of the hexameric ATPase ring of the base takes place on the pre-assembled core particle, and the carboxy-terminal domain of the base Rpt subunits is essential for docking with the core particle. Regulatory-particle assembly is facilitated by three chaperone proteins — Rpn14, Nas6 and Hsm3 — each of which binds to specific Rpt proteins¹ and probably guides interaction between the Rpt proteins and the core-particle α -subunit¹. The chaperone proteins also bind near the carboxy-terminal tail of the Rpt proteins, and the authors propose that they compete with the α -subunit pockets for binding the Rpt tails. In this way, the chaperone proteins may ensure that their partner Rpt binds to the correct α -subunit.

Mutations in the carboxy-terminal domain of the Rpt proteins that reduce their binding to the core particle result in increased amounts of free core particle, but, surprisingly, there is no evidence of free regulatory particle². And although increased amounts of the lid subcomplex are detected, no free base is found². These results suggest that the Rpt proteins are sequestered in intermediate assembly complexes, and Park *et al.*² provide supporting evidence for this by identifying an intermediate subcomplex containing the chaperone

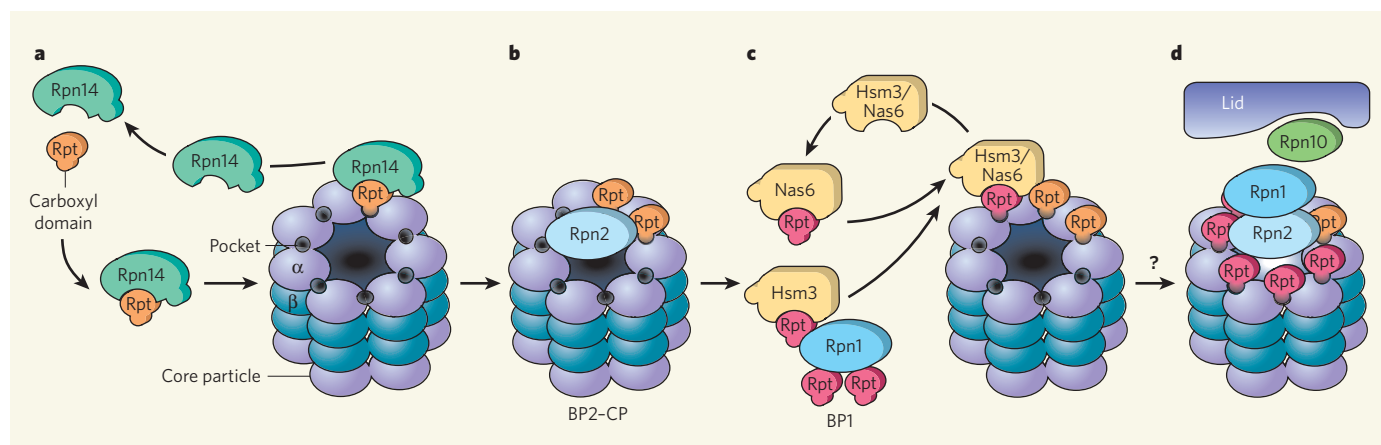


Figure 1 | A model of proteasome regulatory-particle assembly^{1,2}. **a**, ‘Templating’ Rpt4 and Rpt6 proteins (orange) bind the Rpn14 chaperone and connect to the core particle. Rpn14 and the core particle α -subunit compete for binding to the carboxyl-terminal domain of Rpt, and the chaperone is displaced following Rpt- α -subunit interaction. **b**, The non-ATPase subunit Rpn2 might stabilize this complex, termed BP2-CP. **c**, Other Rpt ATPases (pink) are transferred to the core particle by chaperones Hsm3 and Nas6. Hsm3 is found in the intermediate complex BP1, which also contains Rpt1, 2 and 5 and Rpn1. The specific interactions among these proteins and their order of addition to the emerging Rpt ring are not known. For clarity, Rpn1 and Rpn2 are not shown on the core particle. **d**, Completion of the ATPase ring fully activates the proteasome. Formation of the base-core-particle complex is presumably followed by assembly of the intact proteasome. The interaction between the base and the lid is predicted to require Rpn10. Only one regulatory particle is shown.

Hsm3, the ATPases Rpt1, Rpt2 and Rpt5, and the non-ATPase subunit Rpn1, which they call base precursor 1 (BP1).

The order of addition of Rpt proteins to the core particle suggests an assembly line of sequential steps, in which the action of the Rpn14 chaperone precedes those of Hsm3 and Nas6. In this model (Fig. 1), formation of the hexameric ring of the regulatory-particle base is initiated by Rpn14, which anchors its binding partners Rpt4 and Rpt6 on the core particle. The non-ATPase subunit Rpn2 might perform a scaffolding role in stabilizing this complex, named the BP2-core particle (BP2-CP). The Rpn14-dependent step is followed by the delivery of additional ATPase subunits by BP1. The initial placement of Rpt4 and Rpt6 on the core particle might guide subsequent interactions between Rpt subunits and specific binding pockets on the core particle, ensuring that subunits in the emerging ring are correctly aligned, and locking the emerging ring firmly in place. These sequential steps yield an intact base-core-particle complex, which presumably sets the stage for the succeeding assembly of the lid to complete the regulatory particle.

Faulty interaction between the ‘templating’ ATPases (Rpt4 and Rpt6) and the core particle causes a more severe proteasome-assembly defect than mutations in the other ATPases (Rpt1, 2, 3 and 5)². One interpretation of this result is that if Rpt4 or Rpt6 fail to bind to their partner α -subunits in the core particle, the pairwise interactions between the core particle and the ‘late-docking’ ATPases (Rpt1, 2, 3 and 5) could be adversely affected. By contrast, defects in the late-docking ATPases are expected to cause a milder defect in proteasome assembly because preceding core-particle interactions with Rpt4 and Rpt6 would steer them towards interacting with specific α -subunits. A previous report⁶ showed that the non-ATPase subunits Rpn1 and Rpn2 interact with the core

particle and provide a platform for the ATPase ring, suggesting a crucial structural role for these non-ATPase subunits. However, as the carboxyl termini of Rpt subunits can interact directly with intra-subunit pockets in the α -subunit ring¹⁻⁴, the specific role of Rpn1 and Rpn2 in base assembly is unclear.

These findings reiterate the complexity of proteasome assembly, which may allow the generation of proteasomes that vary in their composition and function to satisfy specific cellular needs. Moreover, as free base, intact regulatory particles and core particles can be detected in cell extracts, it is possible that alternative assembly mechanisms exist. Such interactions may be governed by the same chaperone proteins, as these can be co-purified with the base and with free regulatory particle¹. One alternative model might predict base assembly occurring close to the α -subunits, with chaperone release synchronized to complete docking of the ring to the core particle.

As expected, the chaperone proteins are not detected in the mature proteasome — they are presumably released to continue their guiding role in assembly. However, Park *et al.*² find that subtle alterations in the carboxyl-terminal domain of the ATPase subunits allow their partner chaperones to remain bound to the core particle. Interestingly, one report⁷ has proposed that, in some cases, protein degradation by the core particle is followed by disassembly of the regulatory particle, although another study⁸ found that the proteasome remains intact. The role of the chaperone proteins during a cycle of protein degradation is unknown, and it remains to be tested whether the retention of regulatory-particle chaperones on proteasomes (using Rpt mutants) affects disassembly or reassembly. The interaction between regulatory-particle chaperones and the free regulatory particle shown by Roelofs *et al.*¹ suggests that these chaperones are also

required for promoting reassociation between the regulatory particle and the core particle.

Defects in the many human genes that encode proteins involved in the ubiquitin-proteasome pathway have been linked to cancers, such as breast cancer⁹, and neurodegenerative disorders, such as Parkinson’s disease¹⁰. Detailed structural characterization of the core particle led to the development of drugs that inhibit proteasome function, and these drugs have been beneficial in the treatment of a type of blood cancer called multiple myeloma¹¹. However, a major limitation of such drugs is that the proteasome is essential in all cells, and altering its activity affects both normal and abnormal cells. The reports in this issue^{1,2}, and other recent findings¹²⁻¹⁶, show that the requirement for regulatory-particle chaperones in proteasome assembly is evolutionarily conserved, revealing the chaperone proteins as promising targets for drug discovery.

Kiran Madura is in the Department of Biochemistry, Robert Wood Johnson Medical School, Piscataway, New Jersey 08854, USA. e-mail: maduraki@umdnj.edu

- Roelofs, J. *et al.* *Nature* **459**, 861–865 (2009).
- Park, S. *et al.* *Nature* **459**, 866–870 (2009).
- Smith, D. M. *et al.* *Mol. Cell* **27**, 731–744 (2007).
- Gillette, T. G., Kumar, B., Thompson, D., Slaughter, C. A. & DeMartino, G. N. *J. Biol. Chem.* **283**, 31813–31822 (2008).
- Ramos, C. P. & Dohmen, R. *J. Structure* **16**, 1296–1304 (2008).
- Rosenzweig, R., Osmulski, P. A., Gaczynska, M. & Glickman, M. H. *Nature Struct. Mol. Biol.* **15**, 573–580 (2008).
- Babbitt, S. E. *et al.* *Cell* **121**, 553–565 (2005).
- Kriegenberg, F. *et al.* *Cell* **135**, 355–365 (2008).
- Downes, M. & Evans, R. M. *Cancer Cell* **13**, 184–185 (2008).
- Olanow, C. W. & McNaught, K. S. *Mov. Disord.* **21**, 1806–1823 (2006).
- Goldberg, A. L. *Biochem. Soc. Trans.* **35**, 12–17 (2007).
- Le Tallec, B., Barrault, M. B., Guérois, R., Carré, T. & Peyroche, A. *Mol. Cell* **33**, 389–399 (2009).
- Nakamura, Y. *et al.* *Biochem. Biophys. Res. Commun.* **359**, 503–509 (2007).
- Funakoshi, M., Tomko, R. J., Kobayashi, H. & Hochstrasser, M. *Cell* **137**, 887–899 (2009).
- Saeki, Y., Toh-e, A., Kudo, T., Kawamura, H. & Tanaka, K. *Cell* **137**, 900–913 (2009).
- Kaneko, T. *et al.* *Cell* **137**, 914–925 (2009).

REVIEWS

Solar eclipses as an astrophysical laboratory

Jay M. Pasachoff^{1,2}

Observations of the Sun during total eclipses have led to major discoveries, such as the existence of helium (from its spectrum), the high temperature of the corona (though the reason for the high temperature remains controversial), and the role of magnetic fields in injecting energy into—and trapping ionized gases within—stellar atmospheres. A new generation of ground-based eclipse observations reaches spatial, temporal and spectral-resolution domains that are inaccessible from space and therefore complement satellite studies.

Solar eclipses occur when the Moon comes directly between the Earth and the Sun. By a happy coincidence, the Moon and Sun subtend the same angle in the sky to within 10%. As the lunar orbit is tilted by 5° with respect to the plane of the Earth's orbit around the Sun (the ecliptic plane), solar eclipses occur only when both Moon and Sun are sufficiently near the nodes, the crossing points of the Moon's apparent path in the sky and the ecliptic¹. Such events occur at least yearly, though the region on Earth from which the totally eclipsed Sun (Fig. 1) is visible is sufficiently small that travel is usually necessary².

The original motivation for studying solar eclipses was to take advantage of rare opportunities, rather than any specific observational plan based on theory. The nature of the outer solar atmosphere was deduced during the roughly hundred years starting in 1860, when it was definitively associated with the Sun instead of being thought of as a lunar atmosphere. This 'solar corona' appears to people on Earth only during total eclipses (Fig. 2), forming a faint crown (Latin, *corona*) of light that Halley called pearly white. We now know it to be hot and ionized, that is, a plasma.

Because of their dramatic nature, bringing the darkness of twilight even at midday, total solar eclipses have been noted for thousands of

years. A total eclipse results in a darkening by a factor of approximately a million, meaning that the coverage of the last 1% of the everyday solar surface produces a darkening by a remaining factor of about 10,000—so the onset of darkness comes only from a change even smaller than that last 1%. Therefore, it can be deduced, even from ancient chronicles, whether a writer was inside or outside the zone of totality. Such observations have therefore been searched for and found in tablets, parchment, and other writing from Asia, the mid-East, and Europe, and then used to determine variations in the Earth's rotation over thousands of years^{3,4}.

Advantages of contemporary eclipse research

Though many areas of coronal research are well covered from telescopes on satellites, there remain several aspects that are, and remain for the foreseeable future, uniquely able to be studied from Earth at total solar eclipses⁵. The low corona can also be monitored to some extent from mountain-top coronagraphs, such as the K-coronameter of the High Altitude Observatory on Mauna Loa, Hawaii, but even that instrument occults (hides) the lowest corona, and its images are relatively noisy and show little detail⁶. Other coronagraph stations include those at Lomnický štít, Slovakia, and at Sacramento Peak, New Mexico.

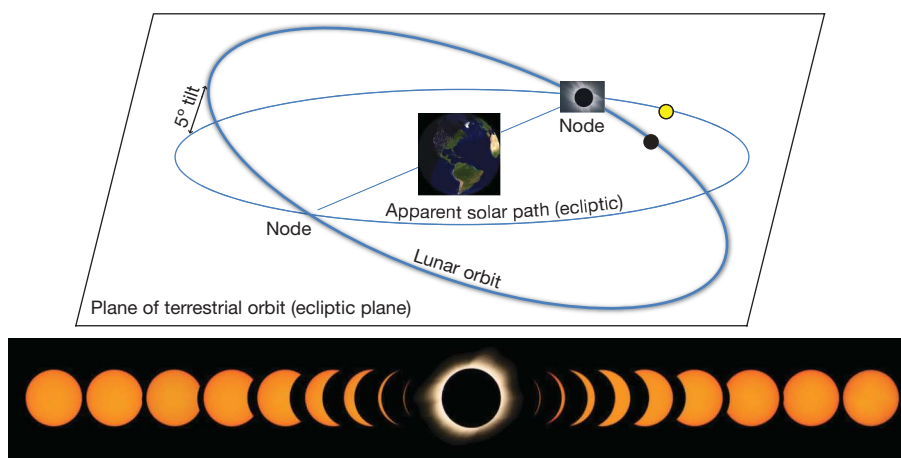


Figure 1 | The geometry of eclipses. Eclipses occur only when both Sun and Moon are sufficiently near to one of the nodes where the Moon's orbit around the Earth crosses the plane of the Earth's orbit around the Sun (the ecliptic plane). The node shifts around the orbit by 18.7° per year, making the eclipse year 5% shorter than a terrestrial year. Sufficiently near the node, a series of partial phases are seen, as shown over a 3-h interval, photographed

through a filter of neutral density 4 or higher. For seconds or minutes, for observers in the umbra of the lunar shadow, the corona can be seen; it is photographed here with no filter, since its total brightness is only roughly that of the full Moon. (eclipse sequence copyright 2006 J.M.P. and R. Ressmeyer, Science Faction).

¹Williams College — Hopkins Observatory, 33 Lab Campus Drive, Williamstown, Massachusetts 01267, USA. ²Department of Planetary Sciences, California Institute of Technology 150-21, Pasadena, California 91125, USA.



Figure 2 | The corona at totality. A composite image of the sunspot-minimum 2008 eclipse corona and prominences, put together from 16 individual frames of a variety of exposure times made at Akademgorodok, Siberia. (Image processing by W. Carlos from exposures by J.M.P., W. G. Wagner and J. Guertin; copyright 2008 W. Carlos and J.M.P.)

Instruments on spacecraft are carefully made for specific purposes, and are locked into their configurations many years in advance of their use. In contrast, eclipse expeditions have the flexibility to use the latest equipment and to take advantage of new theoretical ideas to frame observations. Eclipse expeditions can also take larger solar telescopes than are in space. Further, an eclipse expedition is so much less expensive than any space instrument that, even allowing for equipment and weather failures, total eclipse expeditions are a relatively inexpensive way to obtain a variety of chromospheric and coronal information. Eclipse expeditions also provide a way to test equipment and methods destined for space launch.

A forthcoming eclipse, even the partial phases, can get the attention of local people through (at least) newspaper coverage, and

instructions on safe eclipse observing should be widespread⁷. Unfortunately, there is often a public and even governmental misunderstanding of what an eclipse entails, and a widespread belief that some additional, harmful rays come out of the Sun during an eclipse. All too often, people in general and students of all ages in particular are prevented from seeing the eclipse, sometimes with misinformation about the chance of eye damage that could lead the public to ignore significant public-health warnings about other, more serious matters at later dates⁸.

Eclipse predictions and the Saros

A modern era of eclipse observing may be said to have begun in 1715, with the first predictions in the form of a terrestrial map bearing a delineated path of totality, worked out by and published for Edmond Halley⁹. The observed path was displaced by some tens of kilometres from the prediction, and the revised, actual path was plotted along with predictions for the subsequent European eclipse, that of 1724.

Solar eclipses occur periodically, with eclipses similar in latitude and duration appearing with an 18 year $11\frac{1}{3}$ -day interval¹⁰, with the $\frac{1}{3}$ day allowing the Earth to rotate 120° . The period, known as the Saros, was discovered by the Babylonians but was named by Halley¹¹.

In a given calendar year, there can be as many as five solar eclipses or as few as two. At a given spot on Earth, an eclipse would be viewed only about every 350 years, though that average interval varies with a terrestrial observer's latitude¹². Paths of some successive eclipses overlap, and fortunate viewers at those locations could see two total eclipses with only a one-year interval (Fig. 3)¹³. Now that travel is so convenient, ardent eclipse viewers travel to all total or even to all solar eclipses. An average interval between total eclipses is about 18 months (ref. 14). The next total solar eclipses to cross the United States occur in 2017 and 2024¹⁵.

The shape of the corona, with the streamers and plumes governed by the coronal magnetic field, varies with the solar-activity cycle. The shape of the corona can be accurately predicted in advance with calculations based on the observed surface magnetic field¹⁶. The Ludendorff flattening coefficient measures the overall shape, with sunspot-maximum coronas appearing relatively round because of the large number of streamers at all latitudes, and the sunspot-minimum

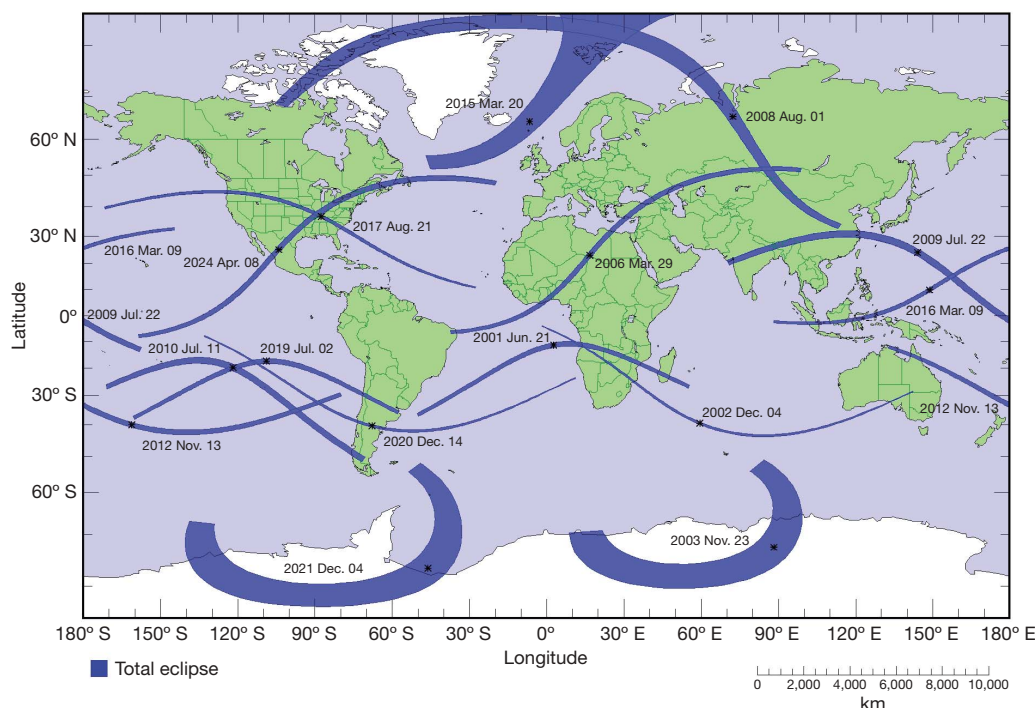


Figure 3 | The paths of total eclipses, 2001–25. Partial phases can be seen to distances of thousands of kilometres to the sides of the umbral paths. (F. Espenak, NASA's Goddard Space Flight Center.)

coronas appearing relatively elongated because mainly equatorial streamers remain (Fig. 3)^{17,18}.

Historical eclipses and corona parameters

Many solar physicists devote their attention to the question of how the solar corona is heated to millions of kelvins, almost 1,000 times hotter than the underlying solar photosphere^{19–21}. The fact of this temperature inversion is a twentieth-century result based on nineteenth-century eclipse observations with the then newly invented spectroscope.

The light from the everyday Sun comes from a layer of semitransparent gas, the photosphere (named from the Greek *photos*, light), at a temperature of about 5,800 K (ref. 22). Observers in the path of totality at a solar eclipse see first a gradual covering of the Sun by the lunar silhouette for over an hour, and then a set of ‘Baily’s beads’, the last bits of the bright photosphere shining through valleys aligned at the edge of the Moon²³. As the Baily’s beads disappear behind the advancing lunar edge (the phenomena occur in reverse at the end of totality), a thin reddish edge, called the chromosphere (named from the Greek *chromos*, colour), appears. Though the reddish hydrogen radiation is most visible to the unaided eye, the chromosphere also radiates thousands of additional spectral lines²⁴. When the chromosphere is covered, as it is within seconds, the corona appears as a halo around the Sun. But what is this corona, what is it made of, and how hot and dense is it? Answering these questions became the goal of eclipse research in the late nineteenth and early twentieth centuries. And the question of how the corona is heated to, as we now know, millions of kelvins, far hotter than the underlying photosphere, remains a major astrophysical question for twenty-first-century astrophysicists.

In the mid-nineteenth century, the near-simultaneous development of spectroscopes that could be taken on expeditions²⁵ and, starting in about 1860, photography²⁶, led to a series of scientifically important eclipse expeditions. Though the motivation was purely observational, in the absence of a theoretical understanding of the Sun, which was thought to resemble the Earth in composition, the developments led decades later to the establishment of solar physics and, eventually, to the rise of astrophysics instead of the traditional astronomy of the time. This traditional nineteenth-century astronomy had a higher quotient of positional measurements—astrometry—and of measuring time-varying phenomena (such as variable-star brightnesses), rather than understanding the causes of stellar evolution and the structure and evolution of the Universe that are centres of today’s astronomical studies.

In 1868, astronomers used their newly available types of instrument to make several expeditions to India to observe the total solar eclipse. P. J. C. Janssen observed the spectrum of the chromosphere that was briefly visible at the beginning and at the end of totality. He saw a bright yellow emission line, apparently at or close to the wavelengths of the known sodium D lines. The spectral lines he saw were so bright that he realized that by broadening his spectrograph’s slit he could see them even without an eclipse, a realization also reached independently soon thereafter by a non-eclipse observer, N. Lockyer (founder of *Nature*) in England. Measured in detail, the position of the bright yellow line was not quite that of sodium D₁ or D₂, so it was called D₃ and was said to come from ‘helium’ (after the Greek Sun god, Helios), since it apparently came from a source only on the Sun. It was not until 1895 that helium was isolated on Earth.

At the eclipses of 1868 and 1869, observed from Siam (now Thailand) and from the United States, respectively, the first coronal emission line was detected²⁷; about 25 years later, it was said to be from the element ‘coronium’, by analogy with the prior discovery of helium. This first ‘coronium’ line is in the green region of the solar spectrum. At subsequent eclipses, a red coronal line was observed, and also several fainter emission lines. At the eclipse of 1870, a more complete set of emission lines from the chromosphere was observed; it is still known as the ‘flash spectrum’ from the way it flashes into view as the Moon completely covers the solar photosphere.

In 1871, not only emission lines but also absorption lines were detected in the coronal spectrum. We now know that the corona we see has three main parts: emission from hot coronal gases, absorption from the solar photospheric spectrum reflected from dust particles in interplanetary space relatively near the Sun, and a continuum from scattering of photospheric light off coronal electrons. These are known as the E-, F-, and K-coronas, respectively. At the same eclipse, spectrographic observations found an emission line detectable almost a solar radius above the limb, a surprising height for the coronal ions that emit the spectral lines (as opposed to the coronal continuum that results from electron scattering).

The solution to the problem of identifying ‘coronium’ did not come until the 1930s. The smooth, absorption-line-free spectrum of the K-corona could indicate that photospheric absorption lines were blurred by exceedingly high electron velocities²⁸. This effect is used to the present day in studying the coronal temperature. Also in the 1930s, with the artificial eclipse formed in the newly invented ‘Lyot coronagraph’, which could observe limited aspects of the corona under special conditions from certain high mountain sites, astronomers found that the coronal green line was 0.9 Å broad; this would have meant high temperature if the atomic mass of the element emitting the line had been known. The solution to identifying the source of the spectral lines came only in 1939, with the extrapolation along isoelectronic sequences from lower states of ionization to discover that the ‘coronium’ lines were actually emitted by highly ionized states of elements such as iron, calcium and nickel²⁹. The coronal green line, for example, came from the ‘forbidden line’ of iron-14, written [Fe XIV], iron stripped of 13 of its 26 electrons (neutral iron is Fe I, so the Roman numerals exceed the ionization number by 1), which would require such a high temperature that the corona must be at millions of kelvins. (The transition probabilities are low for forbidden lines, which are indicated by square brackets, as they do not obey the normal selection rules of spectroscopy.) So all the spectral lines could be explained by extremely ionized states of known elements, and no new element was needed; in any case, the periodic table had been filled in during the decades since the first ‘coronium’ lines had been found. In the current era, spectroscopic observations of coronal lines are made with coronagraphs and at eclipses³⁰.

Though ground-based observations observe the forbidden lines, many of the corresponding coronal permitted lines were discovered on spectra taken by a rocket launched during the 7 March 1970 eclipse, in which totality was convenient for NASA’s rocket-launching range on Wallops Island, Virginia³¹. These observations were taken with a slitless spectrograph, in which no narrow slit is used as the emitting solar crescent of radiation acts as a slit. (The intensity diminishes rapidly with height above the photosphere, so the approximation of a slit is useful but not perfect.) Surprisingly, strong Lyman- α radiation was discovered in the spectra; calculations show that even at coronal temperatures enough neutral hydrogen remains to provide the observed intensity by resonance scattering of Lyman- α radiated in the chromospheric–coronal transition region. The discovery has been exploited by subsequent measurements of coronal parameters such as temperatures, electron densities and ion densities from Lyman- α observations made from rockets, a space shuttle, and satellites³². An independent determination of the high temperature of the corona was made outside eclipse from radio observations³³.

Motivated by the opportunity to observe an eclipse for a time a factor of 10 longer than normally possible, and by the low-scattering background and high infrared atmospheric transmission available at high altitude, a unique eclipse flight was made in a supersonic Concorde in 1973, with 74 min in totality. For an equatorial eclipse near noon, the relative velocity of the lunar umbra on the Earth can dip to 1,730 km h⁻¹ (ref. 34). Several scientific experiments had an unprecedented 74 min of totality^{35,36}. The infrared has also been observed from high-altitude mountain sites, when the 1991 eclipse went over Mauna Kea³⁷. The 2.5-m telescope on the Stratospheric Observatory for Infrared Astronomy (SOFIA), the instrumented

aeroplane soon to begin its scientific work, should be valuable for eclipse studies, since its location above so much of the terrestrial atmosphere's water vapour, along with the extremely low scattering background at its altitude, will allow unprecedented surveying of the coronal infrared spectrum.

Merging eclipse and space observations

Because the intensity gradient of the inner corona is so great, with the intensity diminishing by a factor of $\sim 1,000$ in the first solar radius, space coronagraphs such as that on the Large Angle and Spectrometric Coronagraph (LASCO) on the NASA/ESA Solar and Hemispheric Observatory (SOHO), launched in 1995, are forced by internal scattering to over-occult, hiding not only the solar photosphere but at least 0.4 solar radii above it. Even the more recent coronagraphs on NASA's STEREO (Solar Terrestrial Relations Observatory) mission, launched in 2006, leave such a 'doughnut' of unobserved lower corona extending 0.4 solar radii above the solar limb. At the same time, the corona on the face of the Sun can be observed in the extreme ultraviolet with the Extreme-ultraviolet Imaging Telescope (EIT) on SOHO and a newer pair of similar instruments with the name Extreme Ultraviolet Imager (EUVI) on the two STEREO spacecraft; with NASA's Transition Region and Coronal Explorer (TRACE), to be superseded in late 2009 by NASA's Solar Dynamics Observatory; and in X-rays with the Smithsonian's X-Ray Telescope (XRT) on the Japanese Hinode (English translation, sunrise) spacecraft. The 'doughnut' inaccessible to spacecraft can be filled in on the days of eclipses with ground-based eclipse observations. The combination provides a full view of coronal features from their footpoints on the solar disk (except for the half of the features whose footpoints are on the Sun's far side) through the lower corona, where the solar wind is formed, and onto the streamers in the upper corona observed with LASCO on SOHO or the coronagraphs on STEREO (Fig. 4)³⁸.

The resolution available at eclipses observed from the ground exceeds that available from spacecraft coronagraphs, and spacecraft designers should be trained in part by viewing total solar eclipses. The designer of LASCO said immediately after the 1994 eclipse, and subsequently, that he would have designed his coronagraphs with higher resolution had he seen an eclipse earlier and so realized how much coronal detail is available.

Testing general relativity at eclipses

Total eclipses were in the popular mind in 1919, when the press publicized reports by Arthur Eddington (who was later knighted) that his two eclipse expeditions to Principe, an island off Africa's west equatorial coast, and Sobral, Brazil, confirmed an important prediction by Albert Einstein³⁹. The report galvanized the public and made Einstein into a celebrity, in addition to providing the scientific backing to his general theory of relativity that I describe below, though W. W. Campbell's data reduction from a similar experiment at the 1918 eclipse had contradicted the 1919 result⁴⁰.

Einstein, in 1916, with his general theory of relativity had explained or predicted three astronomical phenomena. The first was the advance of the perihelion of Mercury, that is, the change in orientation of Mercury's elliptical orbit around the Sun by a small amount, only 43 arcsec per century, after the effects of known gravitational perturbations of the planets were subtracted from the observed rate of change. But since this effect was already known observationally, its explanation by Einstein was not a convincing proof of his theory. The second was the deflection of starlight that passes near the Sun, because of the way that the Sun's mass warps space-time. This effect could be tested only at a total solar eclipse, when the Sun would be up but without a blue sky masking the stars. Since this effect had not yet been measured, if Einstein's theory predicted the correct deflection, agreement of observations with theory would provide proof convincing to many, if not most. The third, a gravitational redshift of light leaving the Sun, was not tested until later.

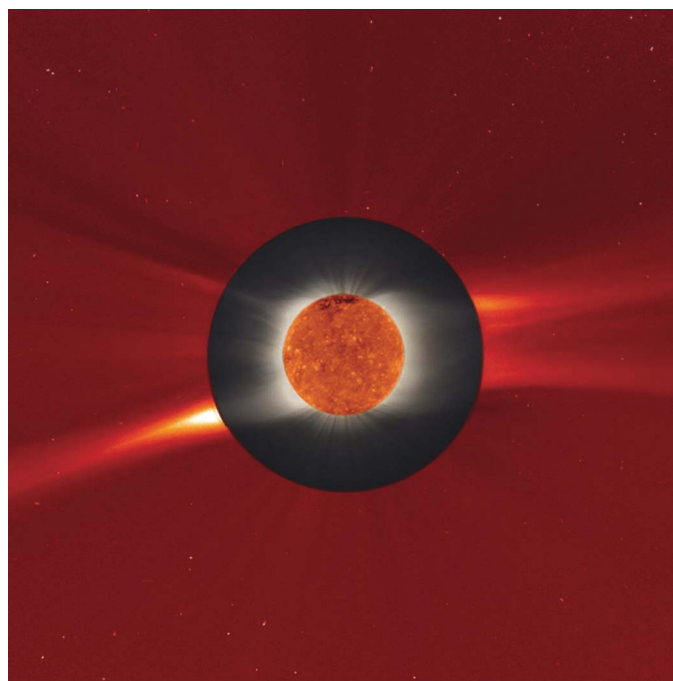


Figure 4 | The 1 August 2008 eclipse from the ground and from space.

Shown is a composite image, consisting of a single eclipse image from 2008 in Siberia (black and white image) sandwiched between an on-disk image from SOHO/EIT in helium radiation at 304 Å in the ultraviolet (central image) and an outer-corona image from SOHO/LASCO. (Eclipse image from Williams College Expedition (J.M.P., B. A. Babcock, M. Baldwin, K. DuPré, M. Freeman, W. G. Wagner, M. Demianski, P. Rosenthal), with A. Nesterenko and I. Nesterenko, State University of Novosibirsk; compositing of eclipse image by W. Carlos. Inner image from the EIT Team, NASA's Goddard Space Flight Center; outer image from the LASCO Team, Naval Research Laboratory. Final composite of eclipse and space images by S. Hill, NASA's Goddard Space Flight Center).

Fortunately for the reputation of Einstein and for posterity, an attempt at the 1914 eclipse to verify an earlier version of Einstein's deflection prediction was ruined when the astronomer who was trying to test the effect by viewing during the totality that crossed Russia was interned there at the outbreak of the First World War; at that time, Einstein's prediction was a factor of 2 lower than his later prediction, and had the eclipse mission succeeded in making its measurement, Einstein's revised calculation might have been perceived as patching rather than pure deduction. In any case, Einstein was later asked what he would have thought if his prediction had not been verified. He famously replied, "Then I would have been sorry for the dear Lord; the theory is correct"⁴¹.

No longer is this Einstein experiment repeated at eclipses, any more than eclipse astronomers search for Vulcan inside Mercury's orbit. The 1919 eclipse had the advantage of the Sun's being in the Hyades star cluster, providing many star images close to the Sun's disk to measure. The 1922 repeat of the experiment with improved equipment provided valuable confirmation that Einstein's prediction was correct⁴². But since that time, the gravitational deflection of starlight by the Sun has been more accurately measured by other methods^{43–46}. The last professional expedition to attempt to verify Einstein's deflection prediction was at the 1973 total solar eclipse; it did not significantly improve prior results^{47,48}.

Heating the corona

There are at least four main types of proposals for coronal heating. (1) Acoustic waves were long thought to provide enough energy, but this class of theory was discarded when spacecraft observations in 1970s did not find enough energy in such waves passing through the

chromosphere. Most of the current generation of theories involves waves on the coronal magnetic field, but whether the waves are (2) Alfvén waves or (3) kink or other magnetic wave types remains controversial⁴⁹. (4) Another set of theories involve ‘nanoflares’, a set of small but frequent explosions on the solar surface⁵⁰. The answer applies not only to the solar corona but also to the coronas of thousands of other stars that are observed with the Chandra X-ray Observatory and other X-ray telescopes. It is not clear what type of observations can distinguish among the theories, so the current state of the art is to make whatever observations are possible in improving time and spatial domains, in order to push the explanatory capabilities of the theories.

One type of eclipse observation that has the potential to distinguish among classes of coronal-heating theory obtains data at much higher cadences (repetition rate) than are available from any solar telescope in space. For example, images of coronal loops have been obtained at 10 Hz, a factor of about 100 faster than can be obtained with NASA’s TRACE, albeit at a resolution of a few arcsec compared with TRACE’s 1 arcsec (740 km) resolution. Recent observations are being studied with Fourier and wavelet techniques, with the aim of distinguishing among theories^{51–53}. Such rapid oscillations are not possible for Alfvén waves on the full coronal loops that are observed with high resolution from spacecraft, but could correspond to, for example, Alfvén waves in surface sheaths on coronal loops. The lack of ability to search for such waves for more than the few minutes available at any given eclipse, not to mention impediments such as atmospheric variations, have prevented definitive measurements from having been made and call out for further eclipse observations in the absence of the ability to make suitable observations from the current or the next generation of solar spacecraft. Relevant observations will be carried out at the 2009 total solar eclipse by several groups.

Detecting fine-scale coronal dynamics

The first known coronal photograph was a daguerreotype taken in Königsberg, Germany, at the eclipse of 28 July 1851. Only the innermost corona showed, but its equatorial elongation is easily seen. A difficulty in photographing the corona is its extreme brightness range, diminishing in intensity by a factor of about 1,000 within one solar radius above the limb. Neither film nor electronic detector can handle such a range in intensities. Two main methods have been used to overcome the problem: (1) a radial filter (Fig. 5a), and (2) post-eclipse compositing (Fig. 5b).

(1) The use of a filter radially graded in intensity, with neutral density above four near the limb (that is, absorption by a factor of 10^4) and zero far out, was pioneered first for the eclipse of 1961 and exploited later by French⁵⁴ and American⁵⁵ astronomers. (2) The methods of post-eclipse compositing of multiple images have improved greatly in recent years⁵⁶. High-contrast imaging reveals wave structure and coronal cavities around limb features like prominences^{57,58}. The resolution is such that motions can be detected in the corona even during the interval that the lunar shadow takes to travel along the path of totality. The dynamics of the corona can be revealed only with such high-resolution observations^{59,60}.

Spectroscopic studies continue, including new observations at current phases of the solar-activity cycle in the visible region of the spectrum, and images in the various coronal emission lines⁶¹. New capabilities of infrared detectors will allow further spectroscopic studies to find additional infrared spectral lines, especially from the altitudes obtainable with the 2.5-m telescope aboard the SOFIA aircraft. High-resolution, Fabry-Perot spectrographs can be and have been taken to eclipses⁶². A spectral passband of 0.016 nm was obtained at the coronal red line at the 2006 eclipse⁶³.

The high-spatial-resolution imaging from post-processing has allowed motions to be measured over the 1 h 9 min period among separated stations along an eclipse path. For example, the motion of a polar plume was measured to be $\sim 70 \text{ km s}^{-1}$ during the 2006 eclipse

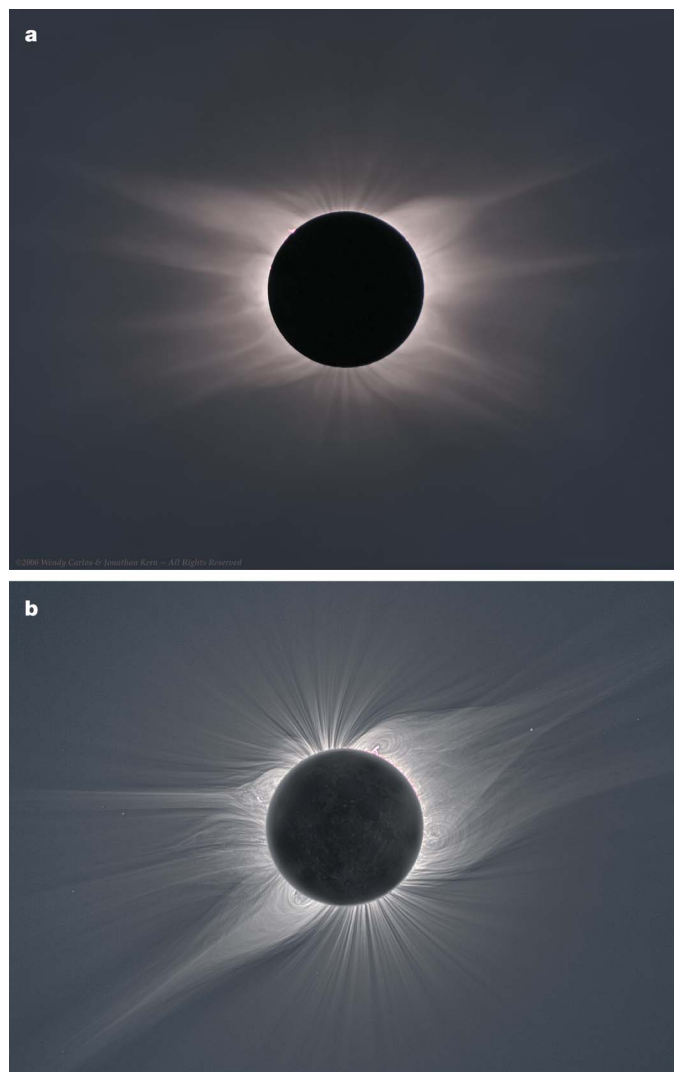


Figure 5 | Correcting for the large dynamic range in coronal intensities. **a**, A composite of several radial-filter images, each also originally showing the same resulting reduction of the wide dynamic range of the corona, taken on film at the 29 March 2006 eclipse from Kastellorizo, Greece. Not only must the radial filter match the average fall-off in intensity to avoid the result showing an obvious ring, but also the alignment of the filter and the solar image must be accurate during the eclipse itself to avoid the equivalent of a Moiré pattern. The radially graded, metal-deposited-on-glass filter declines from neutral density 2.9 at the limb, to 0.7 at 1 solar radius above the limb, to 0.33 at 2 solar radii, to 0.06 at 3 solar radii. The High Altitude Observatory group provided a notable series of radial-filter images at every major eclipse from 1966 to 1984 on film and in 1988 with a CCD. The set of images clearly shows how the shape of the corona varies over the solar-activity cycle. **b**, The total solar eclipse of 1 August 2008, observed from Siberia. Equatorial streamers and polar plumes clearly map the corona’s magnetic field. Earthshine illuminates the silhouetted lunar disk. (Image processing of **a** by W. Carlos from radial-filter images by J. Kern, now of the Observatories of the Carnegie Institution of Washington, as part of the Williams College Expedition, copyright 2006 W. Carlos and J. Kern. Images in **b** from J.M.P., W. G. Wagner and H. Druckmüllerová, image processing by H. Druckmüllerová.)

as the umbra moved from Libya to Egypt to Greece to Turkey, setting a value for comparison with prior plume velocities and deduced lifetimes⁶⁴. Similar measurements are being made from the 2008 eclipse observations by comparing images from Russia, from Mongolia, and, with lower resolution because of the aircraft window, at a latitude of 83° from an aircraft north of Svalbard (midway between Norway and the North Pole)⁶⁵.

Box | Eclipses: public safety and education

The Working Group on Eclipses of the International Astronomical Union (IAU) comprises active eclipse scientists from all over the world, and provides advice to people in countries with forthcoming eclipses. It also works to facilitate and coordinate such matters as duty-free entry of scientific equipment to the relevant countries through which totality passes⁷¹. Additionally, the IAU Commission on Education and Development has a Program Group on public education at the times of eclipses, which includes a professor of optometry in addition to astronomers.

A Solar Eclipse Mailing List (SEML) facilitates correspondence and coordination among amateur and professional eclipse observers. Postings fall into several classes, including plans for future eclipses, results of recent eclipses, calculations of sometimes obscure eclipse statistics, and sightings of eclipses in movies (SEML@yahoogroups.com).

NASA's Goddard Space Flight Center runs a solar eclipse homepage, which has links to maps, charts, and a variety of detailed eclipse publications⁷². Solar eclipse technical publications have recently been prepared for each major eclipse⁷³. They may also be prepared for all the central (total and annular) eclipses in a given year⁷⁴.

The International Year of Astronomy will be marked on 22 July 2009—especially in India, China, and some smaller Japanese islands—by the longest totality in the 18 year $11\frac{1}{3}$ -day Saros cycle, which will be observed by many teams of professionals as well as unprecedentedly large numbers of eclipse enthusiasts, providing an entrée not only for astrophysics but also for public education in all aspects of astronomy and other sciences⁷⁵. It is to be the longest duration of totality that will occur until the year 2132.

Variation of coronal temperature and density

The corona we see during an eclipse is an overlap of the continuous, largely polarized, K-corona, the basically unpolarized F-corona, and in the inner corona an overlay of emission lines. The million-kelvin temperature means that the coronal electrons are moving so rapidly that they blur out the solar photospheric spectrum as scattered in the corona, and the F- and K- coronas can be separated from each other by the residual depth of the absorption lines⁶⁶. At the eclipse of 2001 in Zambia, measurements of the coronal spectrum at several wavelengths were compared with a model to determine the coronal electron temperature and bulk flow speed of the solar wind⁶⁷. The results were extended to test the effects of coronal streamers⁶⁸. At the 2008 total solar eclipse in Siberia, polarization measurements were again obtained in order to separate the coronal components. After the separation of the components, the electron density can be determined⁶⁹. Such determinations that can be made on the Sun only because of its proximity to Earth are extrapolated not only to sunlike stars but also to stars of all types, so our detailed studies at solar eclipses inform our understanding of trillions of stars throughout the Universe.

Few if any other phenomena are available for scientific study only at predictable intervals and then for periods of only seconds or minutes. It is not a surprise, then, that some solar astronomers choose to gather whatever information can be observed at total eclipses even in the absence of plans for definitive astrophysical tests.

Eclipse ecotourism

Seeing an eclipse is so dramatic, with daylight turning abruptly to twilight at midday and unique phenomena becoming briefly visible as totality begins and ends, that even non-scientists have long followed up on eclipse reports and travelled to the path of totality. Eclipse tourism may have begun with the eclipse of 1868, for which the King of Siam made his own, correct, predictions, and led a large team of notables and others to the path of totality. Unfortunately, he was bitten by a malarial mosquito on that expedition and died soon thereafter from that disease.

Major eclipse expeditions, with large quantities of equipment and large numbers of accompanying persons, were common in the

Victorian era, to find out whatever could be found by such observations⁷⁰. In the modern era, the single cruise ship carrying tourists and a few scientists as lecturers into the path of totality off the African coast of the 1973 eclipse has developed into an industry, with several cruise ships and dozens of land expeditions for each totality, as ecotourism grows and people who experienced one eclipse come back for others, bringing along friends to experience the remarkable phenomena.

Future eclipse research

The Moon is receding from the Sun sufficiently slowly that our descendants on Earth will be able to see total eclipses for over 600 million years. In the nearer term, it appears that for decades ground-based capabilities will still allow unique observations to be made from Earth rather than from space. Eventually, propinquity of spacecraft to the Sun plus improvements of space solar telescopes in spatial, temporal and frequency domains may allow such space solar telescopes to take over completely, perhaps adding to observations of distant stellar coronas to explain the coronal-heating problem. At present the paired science and beauty of solar eclipses remain uniquely available to scientists and others in the path of totality.

- Golub, L. & Pasachoff, J. *Nearest Star: The Surprising Science of Our Sun* (Harvard Univ. Press, 2001).
- Held, W. *Eclipses: 2005–2017: A Handbook of Solar and Lunar Eclipses and Other Rare Astronomical Events* (Floris Books, 2005); transl. from *Astronomische Sternstunden* (Verlag Freies Geistesleben, 2005).
- Stephenson, F. R. *Historical Eclipses and Earth's Rotation* (Cambridge Univ. Press, 1997).
- Shows the value that eclipses have for assessing Earth's rotation over thousands of years.
- Vaquero, J. M. *The Sun Recorded Through History: Scientific Data Extracted from Historical Documents* (Springer, 2009).
- Guillermier, P. & Koutchmy, S. *Total Eclipses: Science, Observations, Myths and Legends* (Springer/Praxis, 1999); transl. from *Eclipses Totales: Histoire, Découvertes, Observations* (Masson, 1998).
- Mauna Loa Solar Observatory. (<http://mlso.hao.ucar.edu/>) (2009).
- Pasachoff, J. M. in *Astronomy in the Developing World* (eds Hearnshaw, J. B. & Martinez, P.) 265–268 (Cambridge Univ. Press, 2007).
- Pasachoff, J. M. in *New Trends in Astronomy Teaching* (eds Gougouenheim, L., McNally, D. & Percy, J. R.) 202–204 (IAU Colloquium 162, Cambridge Univ. Press, 1998).
- Pasachoff, J. M. Halley and his maps of the total eclipses of 1715 and 1724. *Astron. Geophys.* **40**, 2.18–2.21 (1999).
- Esenak, F. & Meeus, J. *Five Millennium Catalog of Solar Eclipses: –1999 to +3000 (2000 BCE to 3000 CE)* (Technical Publication TP-2006-214170, NASA, 2008). Provides information about eclipses over an extended period of time, giving both historical and predictive information.
- Cook, A. Halley and the Saros. *Q. J. R. Astron. Soc.* **37**, 349–353 (1996).
- Williams, S. *UK Solar Eclipses from Year 1 (an Anthology of 3,000 Years of Solar Eclipses)* (Clock Tower Press, Leighton Buzzard, UK, 1996).
- Esenak, F. & Meeus, J. *Five Millennium Canon of Solar Eclipses: –1999 to +3000 (2000 BCE to 3000 CE)* (Technical Publication TP-2006-214141, NASA, 2006).
- Littmann, M., Esenak, F. & Willcox, K. *Totality: Eclipses of the Sun* 3rd edn (Oxford Univ. Press, 2008).
- Mobberley, M. *Total Solar Eclipses and How to Observe Them* (Springer, 2007).
- Rušin, V. et al. Comparing eclipse observations of the August 1, 2008 solar corona with an MHD model prediction. *Trans. Am. Geophys. Un.* (Fall Meet.) abstr. SH13B-1524 (2008).
- Rušin, V. in *Last Total Solar Eclipse of the Millennium* (eds Livingston, W. & Özgüç, A.) 17–31 (ASP Conf. Ser. 205, Astronomical Society of the Pacific, 2000).
- Özkan, M. T. et al. in *Modern Solar Facilities—Advanced Solar Science* (eds Kneer, F., Puschmann, K. G. & Wittmann, A. D.) 201–204 (Universitätsverlag Göttingen, 2007); (http://webdoc.sub.gwdg.de/univlag/2007/solar_science_book.pdf).
- Zirin, H. *Astrophysics of the Sun* (Cambridge Univ. Press, 1988).
- Foukal, P. *Solar Astrophysics* 2nd edn (Wiley-VCH, 2004).
- Aschwanden, M. J. *Physics of the Solar Corona: An Introduction with Problems and Solutions* 3rd printing (Praxis, 2009).
- Golub, L. & Pasachoff, J. M. *The Solar Corona* (Cambridge Univ. Press, 1997). Use of not only eclipses but also space observations to provide a survey of current knowledge of our Sun's outer atmosphere.
- Pasachoff, J. M. & Covington, M. *The Cambridge Eclipse Photography Guide* (Cambridge Univ. Press, 1993).
- Menzel, D. H. A study of the solar chromosphere based upon photographs of the flash spectrum taken by Dr. William Wallace Campbell, Director of the Lick Observatory, at the total eclipses of the Sun in 1898, 1900, 1905 and 1908. *Publ. Lick Obs.* **17**, Part 1 1–303 (1931).
- Meadows, A. J. *Early Solar Physics* (Pergamon, 1970).

26. Pasachoff, J. M., Olson, R. J. M. & Hazen, M. L. The earliest comet photographs: Usherwood, Bond, and Donati 1858. *J. Hist. Astron.* **27**, 129–145 (1996).
27. Rayet, G. A. P. Analyse spectrale des protubérances observées pendant l'éclipse totale visible le 18 août 1861, à la presqu'île de Malacca. *C.R. Acad. Sci. LXVII*, 757–759 (1868).
28. Menzel, D. H. & Pasachoff, J. M. On the obliteration of strong Fraunhofer lines by electron scattering in the solar corona. *Publ. Astron. Soc. Pacif.* **80**, 458–461 (1968).
29. Edlén, B. Die Deutung der Emissionslinien im Spektrum der Sonnenkorona. *Z. Astrophys.* **22**, 30–64 (1942).
30. Koutchmy, S., Contesse, L., Viladrich, Ch, Vilinga, J. & Bocchialini, K. in *The Dynamic Sun: Challenges for Theory and Observations* (eds Danesy, D., Poedts, S., de Groof, A. & Andries, J.) 26.1–26.7 (SP-600, ESA, 2005).
31. Gabriel, A. H. et al. Rocket observations of the ultraviolet solar spectrum during the total eclipse of 1970 March 7. *Astrophys. J.* **169**, 595–614 (1971).
32. Kohl, J. L., Gardner, L. D., Strachan, L. & Hassler, D. M. Ultraviolet spectroscopy of the extended solar corona during the SPARTAN 201 mission. *Space Sci. Rev.* **70**, 253–261 (1994).
33. Lang, K. R. *The Sun from Space* (Springer, 2009).
34. Ottewill, G. *The Under-Standing of Eclipses* (Astronomical Workshop, Greenville, South Carolina, 1991).
35. Beckman, J. et al. Eclipse flight of Concorde 001. *Nature* **246**, 72–74 (1973).
36. Léna, P., Hall, D., Soufflot, A. & Viala, Y. The thermal emission of the dust corona, during the eclipse of June 30, 1973. II — Photometric and spectral observations. *Astron. Astrophys.* **37**, 81–86 (1974).
37. Lindsey, C. et al. Extreme-infrared brightness profile of the solar chromosphere obtained during the total eclipse of 1991. *Nature* **458**, 308–310 (1992).
38. SOHO. pick of the week: Solar eclipse, August 1, 2008. (<http://sohowww.nascom.nasa.gov/pickoftheweek/old/01aug2008/>) (2008).
39. Dyson, F. W., Eddington, A. S. & Davidson, C. A determination of the deflection of light by the Sun's gravitational field from observations made at the total eclipse of May 29, 1919. *Phil. Trans. R. Soc. A* **220**, 291–333 (1920).
40. Crellin, J. *Einstein's Jury: The Race to Test Relativity* (Princeton Univ. Press, 2006).
Describes in detail the most famous eclipse in scientific history and its well-known ramifications for testing Einstein's general theory of relativity as well as providing lessons in the philosophy of science.
41. Isaacson, W. *Einstein: His Life and Universe* (Simon & Schuster, 2007).
42. Zirker, J. B. *Total Eclipses of the Sun* (Princeton Univ. Press, 1995).
43. Muhleman, D. O., Ekers, R. D. & Fomalont, E. B. Radio interferometric test of the general relativistic light bending near the sun. *Phys. Rev. Lett.* **24**, 1377–1380 (1970).
44. Shapiro, S. S., Davis, J. L., Lebach, D. E. & Gregory, J. S. Measurement of the solar gravitational deflection of radio waves using geodetic very-long-baseline interferometry data, 1979–1999. *Phys. Rev. Lett.* **92**, 121101 (2004).
45. Froeschle, M., Mignard, F. & Arenou, F. in *Proceedings of the ESA Symposium 'Hipparcos - Venice '97'* (eds Perryman, M. A. C., Bernacca, P. L. & Battrick, B.) 49–52 (SP-402, ESA, 1997).
46. Bertotti, B., Iess, L. & Tortora, P. A test of general relativity using radio links with the Cassini spacecraft. *Nature* **524**, 374–376 (2003).
47. Evans, D. S. & Winget, K. I. *Harlan's Globetrotters: The Story of an Eclipse* (Xlibris, Tucson, 2005).
48. Texas Mauritanian Eclipse Team. Gravitational deflection of light: Solar eclipse of 30 June 1973. I: Description of procedures and results. *Astron. J.* **81**, 452–454 (1976).
49. Van Doorsselaere, T., Nakariakov, V. M. & Verwichte, E. Detection of waves in the solar corona: Kink or Alfvén waves. *Astrophys. J.* **465**, L73–L75 (2008).
50. Klimchuk, J. A. On solving the coronal heating problem. *Sol. Phys.* **234**, 41–77 (2006).
51. Pasachoff, J. M., Babcock, B. A., Russell, K. D. & Seaton, D. B. Short-period waves that heat the corona detected at the 1999 eclipse. *Sol. Phys.* **207**, 241–257 (2002).
52. Williams, D. R. in *SOHO 13—Waves, Oscillations and Small-Scale Transient Events in the Solar Atmosphere: A Joint View from SOHO and TRACE* (ed. Lacoste, H.) 513–518 (SP-547, ESA, 2004).
Comments on measurements pointed toward distinguishing among theories of coronal heating, the major outstanding coronal research problem.
53. Cowdik, R., Singh, J., Saxena, A. K., Srinivasan, R. & Raveendran, A. V. Short-period intensity oscillations in the solar corona observed during the total solar eclipse of 26 February 1998. *Sol. Phys.* **188**, 89–98 (1999).
54. Laffineur, M., Burnichon, M.-L. & Koutchmy, S. Weighted observation of the corona during the total solar eclipse of September 22, 1968. *Nature* **222**, 461–462 (1969).
55. Saito, K. & Tandberg-Hanssen, E. The arch systems, cavities and prominences in the helmet streamer observed at the solar eclipse, November 12, 1966. *Sol. Phys.* **31**, 105–121 (1973).
56. Druckmüller, M., Rušin, V. & Minarovjech, M. A new numerical method of total solar eclipse photography processing. *Contrib. Astron. Obs. Skalnaté Pleso* **36**, 131–148 (2006).
57. Pasachoff, J. M., Rušin, V., Druckmüller, M. & Saniga, M. Fine structures in the white-light solar corona at the 2006 eclipse. *Astrophys. J.* **665**, 824–829 (2007).
Shows the state-of-the-art in coronal imaging and points toward the future of computer processing for improving resolution of eclipse observations.
58. Pasachoff, J. M., Kimmel, S. B., Druckmüller, M., Rušin, V. & Saniga, M. The 8 April 2005 eclipse white-light corona. *Sol. Phys.* **238**, 261–270 (2006).
59. Koutchmy, S. in *Proc. IAU Symposium 233, Multi-Wavelength Investigations of Solar Activity* (eds Stepanov, A. V., Benevolenskaya, E. E. & Kosovichev, A. G.) 509–516 (Cambridge Univ. Press, 2004).
60. Koutchmy, S. et al. CFHT eclipse observations of the very fine-scale solar corona. *Astron. Astrophys.* **281**, 249–257 (1994).
61. Kim, I. S. in *Last Total Solar Eclipse of the Millennium* (eds Livingston, W. & Özgüç, A.) 69–82 (ASP Conf. Ser. 205, Astronomical Society of the Pacific, 2000).
62. Chandrasekhar, T., Ashok, N. M., Rao, B. G., Pasachoff, J. M. & Suer, T.-A. in *Solar Active Regions and 3D Magnetic Structure 273* (26th Meeting of the IAU, Joint Discussion 3, Cambridge Univ. Press, 2006).
63. Noble, M. W. et al. Observing the solar corona with a tunable Fabry-Perot filter. *Appl. Opt.* **47**, 5744–5749 (2008).
64. Pasachoff, J. M. et al. Polar plume brightening during the 29 March 2006 total eclipse. *Astrophys. J.* **682**, 638–643 (2008).
65. Pasachoff, J. M. et al. Coronal observations at the Siberian 2008 total solar eclipse. *Bull. Am. Astron. Soc.* **41**, abstr. 60.03 (2009).
66. Ichimoto, K. et al. Measurement of the coronal electron temperature at the total solar eclipse on 1994 November 3. *Publ. Astron. Soc. Japan* **48**, 545–554, plates 15–16 (1996).
67. Reginald, N. L., St Cyr, O. C., Davila, J. M. & Brosius, J. W. Electron temperature and speed measurements in the low solar corona: Results from the 2001 June eclipse. *Astrophys. J.* **599**, 596–603 (2003).
68. Reginald, N. L., Davila, J. M. & St Cyr, O. C. Effects of streamers on the shape of the K-coronal spectrum. *Sol. Phys.* **225**, 249–265 (2004).
Describes recent use of spectroscopy rather than imaging to study the corona at eclipses.
69. Quémerais, E. & Lamy, P. L. Two-dimensional electron density in the solar corona from inversion of white light images — Application to SOHO/LASCO-C2 observations. *Astron. Astrophys.* **393**, 295–304 (2002).
70. Pang, A. S.-K. *Empire and the Sun: Victorian Solar Eclipse Expeditions* (Stanford Univ. Press, 2002).
71. International Astronomical Union Commission 46 on Education and Development: Program Group on Public Education on the Occasions of Solar Eclipses; Working Group on Solar Eclipses. (<http://www.eclipses.info>).
72. NASA Eclipse Web Site: Solar eclipse page. (<http://eclipse.gsfc.nasa.gov/solar.html>).
73. Espenak, F. & Anderson, J. *Total Solar Eclipse of 2009 July 22* (TP-2006-214141, NASA, 2006).
74. Espenak, F. & Anderson, J. *Annular and Total Solar Eclipses of 2010* (TP-2008-214171, NASA, 2008).
75. Demircan, O., Selam, S. O. & Albayrak, B. *Solar and Stellar Physics through Eclipses* (ASP Conf. Ser. 370, Astronomical Society of the Pacific, 2007).

Acknowledgements The author is glad to acknowledge his mentors and eclipse colleagues over the 50-year period that he has been observing his 48 solar eclipses, including D. H. Menzel, J. P. Schierer, B. A. Babcock, S. P. Souza, V. Rušin, M. Druckmüller, M. Saniga and others; L. Golub, R. W. Noyes, E. H. Avrett and H. Zirin for other aspects of solar physics; F. Espenak for eclipse-path calculations; and generations of undergraduate students from Williams College who have participated in expeditions and/or studied eclipse data. At various times, his eclipse research has been supported by the Committee for Research and Exploration of the National Geographic Society, the Astronomy and Atmospheric Sciences Divisions of the National Science Foundation, and the Heliophysics Division of NASA, as well as by Williams College. He thanks M. Brown and the Division of Geological and Planetary Sciences of the California Institute of Technology for hospitality during the writing of this Review. The preparation of the article benefited from NASA's Astrophysical Data System.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence should be addressed to J.M.P. (jmp@williams.edu).

Early Cambrian ocean anoxia in South China

Arising from: M. Wille, T. F. Nägler, B. Lehmann, S. Schröder & J. D. Kramers *Nature* **453**, 767–769 (2008)

The cause of the most marked changes in the evolution of life, which define the first-order stratigraphic boundary between the Precambrian and the Phanerozoic eon, remains enigmatic and a highly topical subject of debate. A global ocean anoxic event, triggered by large-scale hydrogen sulphide (H_2S) release to surface waters, has been suggested by Wille *et al.*¹, on the basis of two data sets from South China and Oman, to explain the fundamental biological changes across the Precambrian/Cambrian (PC/C) boundary. Here we report a new precise SHRIMP U–Pb zircon age of 532.3 ± 0.7 million years (Myr) ago (Fig. 1) for a volcanic ash bed in the critical unit that reflects the ocean anoxic event, the lowermost black shale sequence of the Niutitang Formation in the Guizhou Province, South China. This age is significantly younger than the precise PC/C boundary age of 542.0 ± 0.3 Myr ago², approximately 10 Myr younger than the extinction of the Ediacaran fauna, and thus challenging the view of a major ocean anoxic event having been responsible for the major changes in the direction of evolution at the PC/C boundary.

So far no reliable precise radiometric data have been available to constrain the age of the widespread Lower Cambrian black shale series (extending over some 1,600 km in a northeasterly direction along the Yangtze Platform) above the PC/C boundary in South China. Wille *et al.*¹ used an imprecise Re–Os age (541 ± 16 Myr ago) of a distinct Ni–Mo sulphide marker bed in the lowermost part of the Niutitang Formation black shale sequence as a basis for their suggestion that a global “sulphide poisoning” caused the drastic changes of life and its distribution across the PC/C boundary. Our new precise SHRIMP U–Pb zircon age data (Fig. 1) from a volcanic ash bed just below the Ni–Mo sulphide marker in the lowermost Niutitang Formation now suggests that the Ni–Mo sulphide precipitation cannot have occurred before 532.3 ± 0.7 Myr ago. By then, the first stage of the ‘Cambrian explosion’—documented by the rapidly expanding trace and body fossil record and a greatly diversifying multiphyletic small shelly fauna (summarized by ref. 3)—was well underway.

A global transient signal was proposed by Wille *et al.*¹ by jointly plotting Mo isotope data from the Ni–Mo sulphide bed, positioned

several metres above the PC/C boundary, with data from shale samples collected from other localities in South China (Ganziping and Yuanling) and in Oman, assuming that all of these samples are contemporaneous (Fig. 1 of ref. 1). Our new age data show that this assumption is not justified because stratigraphic thickness above the PC/C boundary marker does not translate linearly into a single time-scale for these locations. The sections studied in South China include a prolonged disconformity, overlain by a diachronous transgression of the basal Niutitang and Zhujiang Formations (Guizhou and Yunnan Provinces, respectively). At the PC/C section studied by us in the Guizhou Province, most of the 10-Myr time span between the top of the Precambrian Dengying Formation and the dated volcanic ash bed appears to be represented by a hiatus at the lithostratigraphic boundary. The new age data therefore offer a reliable constraint on the maximum duration of non-sedimentation at the PC/C boundary (that is, approximately 10 Myr), which is widespread on the Yangtze Platform⁴.

Although upwelling of H_2S -rich deep ocean water may still be a good explanation for the widespread black shale deposits and their unusual metal contents, a sea-floor hydrothermal venting event also remains a viable alternative in the Chinese sections^{5–7}. In conclusion, the new age data show that the assumption made by Wille *et al.*¹ that the Mo isotope signals in the sections from Oman and South China are contemporaneous cannot be upheld, hence their global significance must be questioned.

METHODS

Sample ZN0909 was collected in 2007 from a volcanic ash bed in a road-cut approximately 200 m east of the Zhongnancun section ($27^\circ 41' 24.6''$ N, $106^\circ 40' 45.2''$ E). Zircons were extracted from approximately 10 kg of rock sample, mounted, polished and examined. The prismatic grains range in length from 70 to 150 μm . Most of them are euhedral, colourless, transparent and cracked to various degrees. We measured absolute abundances of U and Pb and their isotopic ratios of zircons on the SHRIMP II ion microprobe at Curtin University, Perth, through the Remote Operation System at the Beijing SHRIMP Center.

Shao-Yong Jiang^{1,2}, **Dao-Hui Pi**^{1,3}, **Christoph Heubeck**⁴, **Hartwig Frimmel**⁵, **Yu-Ping Liu**³, **Hai-Lin Deng**³, **Hong-Fei Ling**¹ & **Jing-Hong Yang**¹

¹State Key Laboratory for Mineral Deposits Research, Department of Earth Sciences, Nanjing University, Nanjing 210093, China.

e-mail: shyjiang@nju.edu.cn

²Beijing SHRIMP Center, Chinese Academy of Geological Sciences, Beijing 100037, China.

³State Key Laboratory of Ore Deposit Geochemistry, Institute of Geochemistry, Chinese Academy of Sciences, Guiyang 550002, China.

⁴Institut für Geologische Wissenschaften, Freie Universität Berlin, D-12249 Berlin, Germany.

⁵Geodynamics & Geomaterials Research Division, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany.

Received 3 January; accepted 23 March 2009.

- Wille, M., Nägler, T. F., Lehmann, B., Schröder, S. & Kramers, J. D. Hydrogen sulphide release to surface waters at the Precambrian/Cambrian boundary. *Nature* **453**, 767–769 (2008).
- Amthor, J. E. *et al.* Extinction of *Cloudina* and *Namacalathus* at the Precambrian–Cambrian boundary in Oman. *Geology* **31**, 431–434 (2003).
- Li, G. *et al.* Early Cambrian metazoan fossil record of South China: Generic diversity and radiation patterns. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **254**, 229–249 (2007).
- Zhu, M. *et al.* Sinian and early Cambrian stratigraphic framework for shallow- to deep-water environments of the Yangtze Platform: an integrated approach. *Prog. Nat. Sci.* **13**, 951–960 (2003).
- Steiner, M., Willis, E., Erdtmann, B. D., Zhao, Y. L. & Yang, R. D. Submarine-hydrothermal exhalative ore layers in black shales from South China and associated

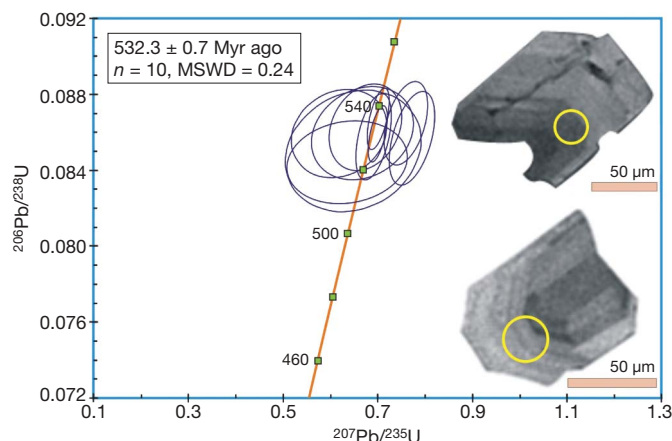


Figure 1 | U–Pb concordia plot for zircon grains from a volcanic ash bed in the lowermost black shale sequence of the Niutitang Formation in Guizhou Province, South China. The analysed spots shown in yellow circles and the zircons display a clear oscillatory magmatic zonation (see inserted cathodoluminescence images) without inherited cores and were therefore chosen for constraining the age of crystallization (that is, 532.3 ± 0.7 Myr ago; with 1 SE error and MSWD (mean square of weighted deviates) = 0.24).

- fossils—insights into a Lower Cambrian facies and bio-evolution. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **169**, 165–191 (2001).
6. Jiang, S. Y. *et al.* Trace- and rare-earth element geochemistry and Pb–Pb dating of black shales and intercalated Ni–Mo–PGE–Au sulfide ores in Lower Cambrian strata, Yangtze Platform, South China. *Mineralium Deposita* **41**, 453–467 (2006).

7. Jiang, S. Y. *et al.* Extreme enrichment of polymetallic Ni–Mo–PGE–Au in Lower Cambrian black shales of South China: an Os isotope and PGE geochemical investigation. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **254**, 217–228 (2007).

doi:10.1038/nature08048

Wille *et al.* reply

Replying to: S.-Y. Jiang *et al.* *Nature* **459**, doi:10.1038/nature08048 (2009)

Jiang *et al.*¹ present a new SHRIMP U–Pb zircon age of 532.3 ± 0.7 million years (Myr) ago for an ash bed in the lowermost black shale sequence of the Niutitang Formation, China, and claim that the data presented in our recent paper² do not firmly support the idea that the biological and environmental changes at the Precambrian/Cambrian transition can be explained by a single global hydrogen sulphide (H₂S) release event. Their new age seems to be supported by another recent SHRIMP investigation³ which indeed suggests that the Chinese metal-enriched sulphide layer does not represent the Precambrian/Cambrian boundary and shows that the redox history of both basins (Oman and South China) was much more complex.

However, as outlined previously⁴, SHRIMP data are not always reliable when it comes to solving stratigraphical problems. For example, with regards to the Yangtze platform, SHRIMP data on the lowermost Doushantuo Formation gave an age of 621 ± 7 Myr ago⁵, whereas single zircon U–Pb data gave an age of 632.5 ± 0.5 Myr ago⁶. This offset is very similar in value and direction to the offset between the Precambrian/Cambrian boundary and the SHRIMP age of Jiang *et al.*¹. If indeed this is related to a systematic inaccuracy of the applied SHRIMP technique, it seems inevitable that single zircon analyses must be undertaken to solve the age problem in the case under discussion.

New age data⁷ confirm an age of 542–541 Myr ago for the base of the Ara A4 cycle, and point to a stratigraphic time interval for the measured section of ~1 Myr. This is in good agreement with the time generated by our model, in favour of an ocean overturn in the lowermost Cambrian. If the age of 532.3 ± 0.7 Myr ago in China can be confirmed, the Mo isotopic signals from Oman and South China would be of diachronous origin. They would thus reflect two successive events, each one less resolved than the initially suggested combined profile. This would affect the temporal resolution and render the quantifications of the model of ref. 2 more uncertain, but would not affect the model's main qualitative conclusion. The most plausible explanation for the Mo isotopic excursions observed in both basins in combination with the same overall Mo isotopic composition is a transient Mo seawater signal caused by excess H₂S.

Jiang *et al.*¹ seem to suggest that the unusual metal enrichment in the Chinese black shales may not be related to scavenging from sea water in a euxinic environment, but from sea-floor hydrothermal venting. However, our Mo isotope data exclude a hydrothermal origin

of molybdenum, which is the element most enriched in these rocks, together with a broad suite of other redox-dependent metals⁸. The consistent $\delta^{98/95}\text{Mo}$ value of $1.1 \pm 0.1\%$ in the basal sulphide marker bed defines the Mo isotope composition of the sea water at that stratigraphic time interval, and the subsequent oscillation of Mo isotope composition in the overlying shale attests to a transitional seawater signal. It remains to be tested in other stratigraphic sections if this oscillation is indeed a global signal, possibly over an extended 10-Myr time interval in the Early Cambrian.

Martin Wille¹, Thomas F. Nägler², Bernd Lehmann³, Stefan Schröder^{4†} & Jan D. Kramers²

¹Research School of Earth Sciences, Australian National University, 0200 Canberra, Australia.

e-mail: martin.wille@anu.edu.au

²Institute for Geological Sciences, University of Bern, Baltzerstrasse 3, 3012 Bern, Switzerland.

³Institute of Mineralogy and Mineral Resources, Technical University of Clausthal, 38678 Clausthal-Zellerfeld, Germany.

⁴Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

[†]Present address: Total E&P, Ave Larribau, F-64018 Pau, France.

- Jiang, S.-Y. *et al.* Early Cambrian ocean anoxia in South China. *Nature* **459**, doi:10.1038/nature08048 (2009).
- Wille, M., Nägler, T. F., Lehmann, B., Schröder, S. & Kramers, J. D. Hydrogen sulphide release to surface waters at the Precambrian/Cambrian boundary. *Nature* **453**, 767–769 (2008).
- Chen, D., Wang, J., Qing, H., Yan, D. & Li, R. Hydrothermal venting activities in the Early Cambrian, South China: Petrological, geochronological and stable isotope constraints. *Chem. Geol.* **258**, 168–181 (2009).
- Bowring, S. A., Grotzinger, J. P., Condon, D. J., Ramezani, J. & Newall, M. J. Reply to comment: Oman Chronostratigraphy. *Am. J. Sci.* **309**, 91–96 (2009).
- Zhang, S. H. *et al.* U–Pb sensitive high-resolution ion microprobe ages from the Doushantuo Formation in south China: Constraints on late Neoproterozoic glaciations. *Geology* **33**, 473–476 (2005).
- Condon, D. *et al.* U–Pb ages from the Neoproterozoic Doushantuo Formation, China. *Science* **308**, 95–98 (2005).
- Bowring, S. A. *et al.* Geochronologic constraints on the chronostratigraphic framework of the Neoproterozoic Huqf Supergroup, Sultanate of Oman. *Am. J. Sci.* **307**, 1097–1145 (2007).
- Lehmann, B. *et al.* Highly metalliferous carbonaceous shale and Early Cambrian seawater. *Geology* **35**, 403–406 (2007).

doi:10.1038/nature08049

ARTICLES

Dual nature of the adaptive immune system in lampreys

Peng Guo^{1*}, Masayuki Hirano^{1*}, Brantley R. Herrin^{1*}, Jianxu Li¹, Cuiling Yu¹, Andrea Sadlonova¹ & Max D. Cooper¹

Jawless vertebrates use variable lymphocyte receptors (VLR) comprised of leucine-rich-repeat (LRR) segments as counterparts of the immunoglobulin-based receptors that jawed vertebrates use for antigen recognition. Highly diverse VLR genes are somatically assembled by the insertion of variable LRR sequences into incomplete germline VLRA and VLRB genes. Here we show that in sea lampreys (*Petromyzon marinus*) VLRA and VLRB anticipatory receptors are expressed by separate lymphocyte populations by monoallelic VLRA or VLRB assembly, together with expression of cytosine deaminase 1 (CDA1) or 2 (CDA2), respectively. Distinctive gene expression profiles for VLRA⁺ and VLRB⁺ lymphocytes resemble those of mammalian T and B cells. Although both the VLRA and the VLRB cells proliferate in response to antigenic stimulation, only the VLRB lymphocytes bind native antigens and differentiate into VLR antibody-secreting cells. Conversely, VLRA lymphocytes respond preferentially to a classical T-cell mitogen and upregulate the expression of the pro-inflammatory cytokine genes interleukin-17 (IL-17) and macrophage migration inhibitory factor (MIF). The finding of T-like and B-like lymphocytes in lampreys offers new insight into the evolution of adaptive immunity.

The ability to mount specific immune responses to a virtually unlimited variety of antigens is apparently unique to vertebrates. The adaptive immune system in jawed vertebrates is characterized by antigen-specific cellular and humoral responses, which together provide durable protective immunity¹. This division of labour is conducted by developmentally separate, but functionally intertwined lineages of clonally diverse lymphocytes, named T and B cells because they are generated in the thymus or the avian bursa of Fabricius and mammalian bone marrow^{2,3}. For antigen recognition, both populations use the same type of immunoglobulin-based receptors. The T and B cell receptors (TCR and BCR, respectively) are assembled during lymphocyte differentiation by the recombination of different variable (V), diversity (D) and joining (J) immunoglobulin gene segments, imprecise V(D)J splicing, and insertion of non-templated nucleotides at the junctions^{4–7}. The assembly process is usually completed only on one allele, so that each T and B cell expresses a different antigen recognition receptor as the basis for clonal diversity⁸. Although the TCR and BCR genes are remarkably similar, the antigen-binding repertoires of T and B cells are very different, primarily because immature T cells undergo clonal selection on the basis of how well they recognize major histocompatibility complex (MHC) class I or class II molecules and their self-peptide cargos^{9–12}. This intrathymic selection serves to eliminate T cells that can recognize and damage self tissues. The ensuing collaboration between T and B cells to achieve effective cellular and humoral immunity involves direct cell contact and indirect communication via cytokines and their receptors¹.

When and how this complex adaptive immune system arose are subjects of much conjecture. Even the most basal jawed vertebrates, the cartilaginous sharks, skates and rays, have a thymus in which T-cell progenitors assemble their TCR genes, and haematopoietic tissues in which B cell progenitors assemble their BCR genes^{13,14}. Although the extant jawless vertebrates lampreys and hagfish were noted previously to have lymphocyte-like cells and to produce antigen-specific agglutinins after immunization, neither a recognizable thymus, nor TCR, immunoglobulin or MHC genes have been found^{15–18}. More recent studies indicate that these agnathans generate clonally diverse anticipatory receptors by

the assembly of genes for VLRs comprised of LRR subunits and an invariant membrane-proximal stalk region¹⁹. The two germline VLR genes, VLRA and VLRB, are incomplete in that they have coding sequences only for portions of the amino- and carboxy-terminal LRR subunits and for the stalk region^{19–23}. However, each germline VLR gene is flanked by hundreds of LRR-encoding sequences, and these are randomly used as templates to add the missing LRR segments needed for completion of a mature VLR gene²². A gene conversion mechanism has been postulated for VLR assembly^{21,23,24}, and candidate activation-induced deaminase (AID)/APOBEC participants have been identified in the lamprey²¹. VLRB assembly has been shown to be confined to one allele^{19,23}, ensuring that each VLRB lymphocyte expresses a receptor of unique sequence, but it is not yet known whether this is also true for VLRA.

In earlier studies, we found that the VLRB-bearing lymphocytes resemble B lymphocytes of jawed vertebrates. They can bind bacterial or erythrocyte antigens and respond by proliferation and differentiation into plasmacytes that secrete multimeric VLRB antibodies specific for protein or carbohydrate epitopes^{25,26}. In this study, we used VLRA-specific antibodies to identify a second population of lamprey lymphocytes and to determine their VLR gene assembly status, gene expression profile, tissue distribution, morphology and functional responses to antigens and mitogens in comparison with the VLRB lymphocytes. Our findings indicate that VLRA lymphocytes are surprisingly similar to the T cells in jawed vertebrates, and suggest functional interactions between the VLRA and VLRB lymphocyte populations.

Identification of VLRA⁺ and VLRB⁺ lymphocyte populations

The invariant stalk regions of the lamprey VLRA and VLRB proteins share less than 11% amino acid sequence identity, a difference that allowed the production of VLRB-specific monoclonal antibodies for use in characterizing the VLRB-bearing lymphocytes^{22,25}. To identify VLRA-bearing lymphocytes, we used a VLRA-immunoglobulin Fc fusion protein as an immunogen to produce two rabbit antisera and three mouse monoclonal antibodies. The VLRA specificity of the antibody preparations was established by comparative analysis with a

¹Emory Vaccine Center and Department of Pathology and Laboratory Medicine, Emory University, 1462 Clifton Road North-East, Atlanta, Georgia 30322, USA.

*These authors contributed equally to this work.

VLRB monoclonal antibody in ELISA and immunoblot assays (Supplementary Fig. 1). Using these VLR-specific antibodies for immunofluorescence staining assays, we could identify discrete VLRA⁺ and VLRB⁺ lymphocyte populations in the principal lymphoid tissues of lamprey larvae: the blood, kidneys, typhlosole and gill region (Fig. 1a). VLRA-bearing lymphocytes were detected in the same relative frequencies by each of the VLRA-specific reagents (Supplementary Fig. 2). Notably, the VLRB⁺ lymphocytes significantly outnumbered VLRA⁺ lymphocytes in all tissues, except the gill region. The VLRB predominance was most marked in the blood and kidneys, in which the VLRB:VLRA ratio was approximately 8:1 (Fig. 1b); examination of adult blood samples indicated persistent VLRB predominance. These findings indicate that VLRA- and VLRB-bearing cells belong to separate lymphocyte populations that are dispersed throughout the body in a characteristic tissue distribution pattern.

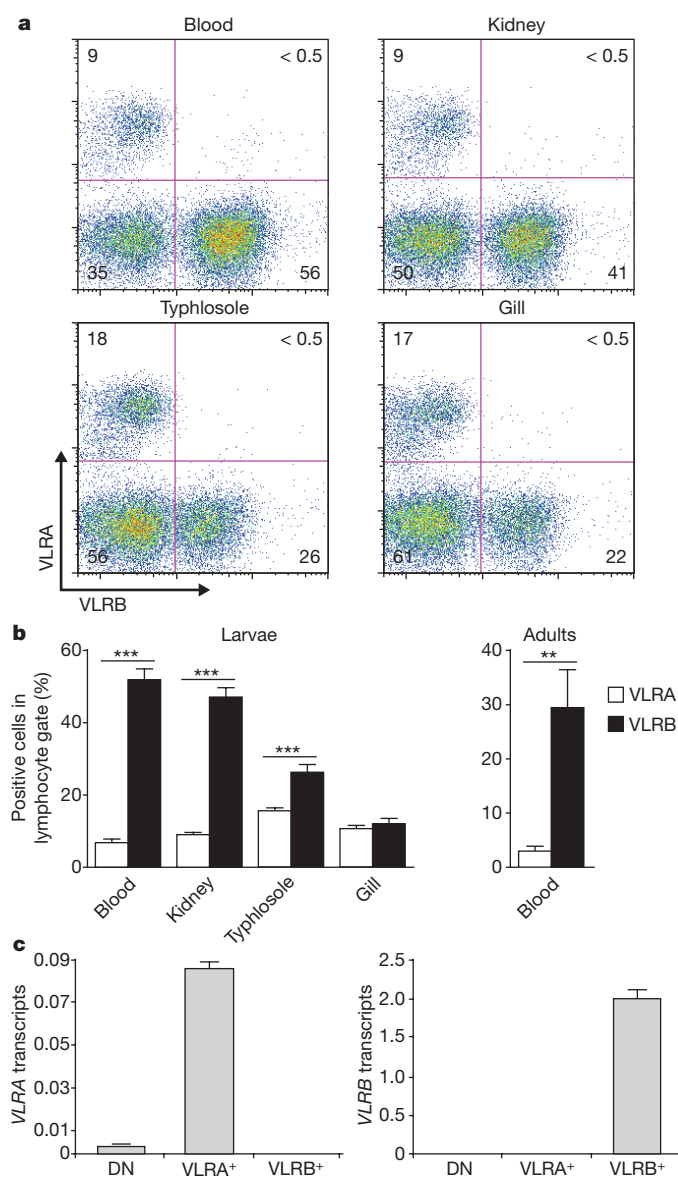


Figure 1 | VLRA and VLRB expression define distinct lymphocyte populations. **a**, FACS analysis of lymphocyte-gated cells stained with anti-VLRA (R110) and anti-VLRB (4C4) antibodies. **b**, Percentages of VLRA- and VLRB-expressing lymphocytes in lamprey tissues. Larvae (left) $n = 23$, except gill $n = 6$; adults (right) $n = 8$; *** $P < 0.01$ and *** $P < 0.001$. **c**, Lymphocyte-gated cells were stained as in **a** and separated on the basis of surface expression of VLRA and VLRB by FACS. Sorted cells were analysed for VLRA (left) and VLRB (right) transcripts by quantitative real-time PCR, $n = 3$. Error bars indicate s.e.m.

Genetic basis of VLRA versus VLRB expression

To explore the genetic basis for the differential expression of VLR types by the two lymphocyte populations, we used fluorescence-activated cell sorting (FACS) to purify VLRA⁺, VLRB⁺ and VLRA⁻ VLRB⁻ double-negative populations of cells having the light scatter characteristics of lymphocytes. The VLRA⁺ and VLRB⁺ lymphocytes could be separated cleanly (>95% purity), whereas the double-negative population included several cell types, most of which were nucleated thrombocytes. When VLR transcription was evaluated for each population by quantitative PCR (qPCR) analysis, the VLRA⁺ lymphocytes expressed VLRA transcripts exclusively, VLRB⁺ lymphocytes expressed VLRB transcripts exclusively, and the double-negative population expressed few or no VLR transcripts of either type. A representative experiment illustrating these results for adult blood cells is shown in Fig. 1c, and the same pattern of selective VLRA and VLRB expression was observed for the sorted populations of blood, kidney and typhlosole cells from larval stage lampreys 3–4 years of age (Supplementary Fig. 3).

When VLRA and VLRB gene configurations were examined for the VLRA⁺, VLRB⁺ and double-negative cells using paired 5' and 3' PCR primers (Fig. 2a, b), we observed mature VLRA assembly only in the VLRA⁺ lymphocyte population, and the same relationship held for VLRB assembly by VLRB⁺ lymphocytes. Mature and germline VLRA PCR products were observed in equal proportions (Supplementary Fig. 4), indicating that VLRA assembly occurs on one allele, a finding that parallels the monoallelic pattern of VLRB assembly demonstrated previously^{19,23}. Because the lamprey AID-like deaminases have been postulated to function in the gene conversion mechanism postulated for VLR gene assembly²¹, we examined the expression of CDA1 and CDA2 in the sorted cell populations. Notably, CDA1 expression was observed only in VLRA⁺ cells, whereas CDA2 expression was limited to the VLRB⁺ population (Fig. 2c). These findings suggest that the monoallelic VLRA versus VLRB gene assembly in the two lymphocyte populations is linked to expression of the genes for the CDA1 or CDA2 deaminases, respectively.

Distinctive gene expression profiles

Examination of the VLRA⁺, VLRB⁺ and double-negative populations for expression of a selected panel of genes indicated that they have different gene expression profiles, and the distinctions were particularly informative for the two lymphocyte populations. The discriminating genes expressed by VLRA and VLRB lymphocyte populations included genes for transcription factors, Notch1, cytokine/chemokines and their receptors, Toll-like receptors (TLRs), and various signalling molecules (Fig. 3, Supplementary Tables 1–3 and Supplementary Fig. 5). A surprising number of the genes that were selectively expressed by VLRA lymphocytes are orthologues of genes that are typically, although not exclusively, expressed by T lineage cells in jawed vertebrates. The genes expressed preferentially by VLRA lymphocytes included: several transcription factors that may be used for T-cell differentiation such as GATA2/3 (refs 27, 28), c-Rel²⁹, aryl hydrocarbon receptor (AHR)³⁰ and BCL11b³¹, the CCR9 chemokine receptor involved in lymphocyte progenitor homing to the thymus³², the T-cell fate-determining molecule Notch1 (ref. 33), the tyrosine phosphatase receptor protein CD45 that is essential for T-cell development³⁴, the IL-8 receptor CXCR2 (ref. 35), and two pro-inflammatory cytokines made by T cells—IL-17 (refs 36, 37) and MIP^{38,39}. The contrasting gene expression profile noted for the VLRB population included: the gene for the haematopoietic progenitor homing receptor CXCR4 (ref. 40), the herpesvirus entry mediator/tumour necrosis factor receptor superfamily, member 14 (TNFRSF14) that binds to LIGHT on T cells⁴¹, two components of the BCR-mediated signalling cascade, Syk⁴² and the B cell adaptor protein (BCAP)⁴³, the immunoglobulin superfamily member TCRL that has inhibitory potential⁴⁴, the chemotactic inflammatory cytokine IL-8 (ref. 45), the IL-17 receptor³⁶, and the Toll-like receptor orthologues TLR2abc, TLR7 and TLR10, the ligation of which is important for B cell activation^{46–48}. This analysis, albeit limited by the incomplete status of the lamprey

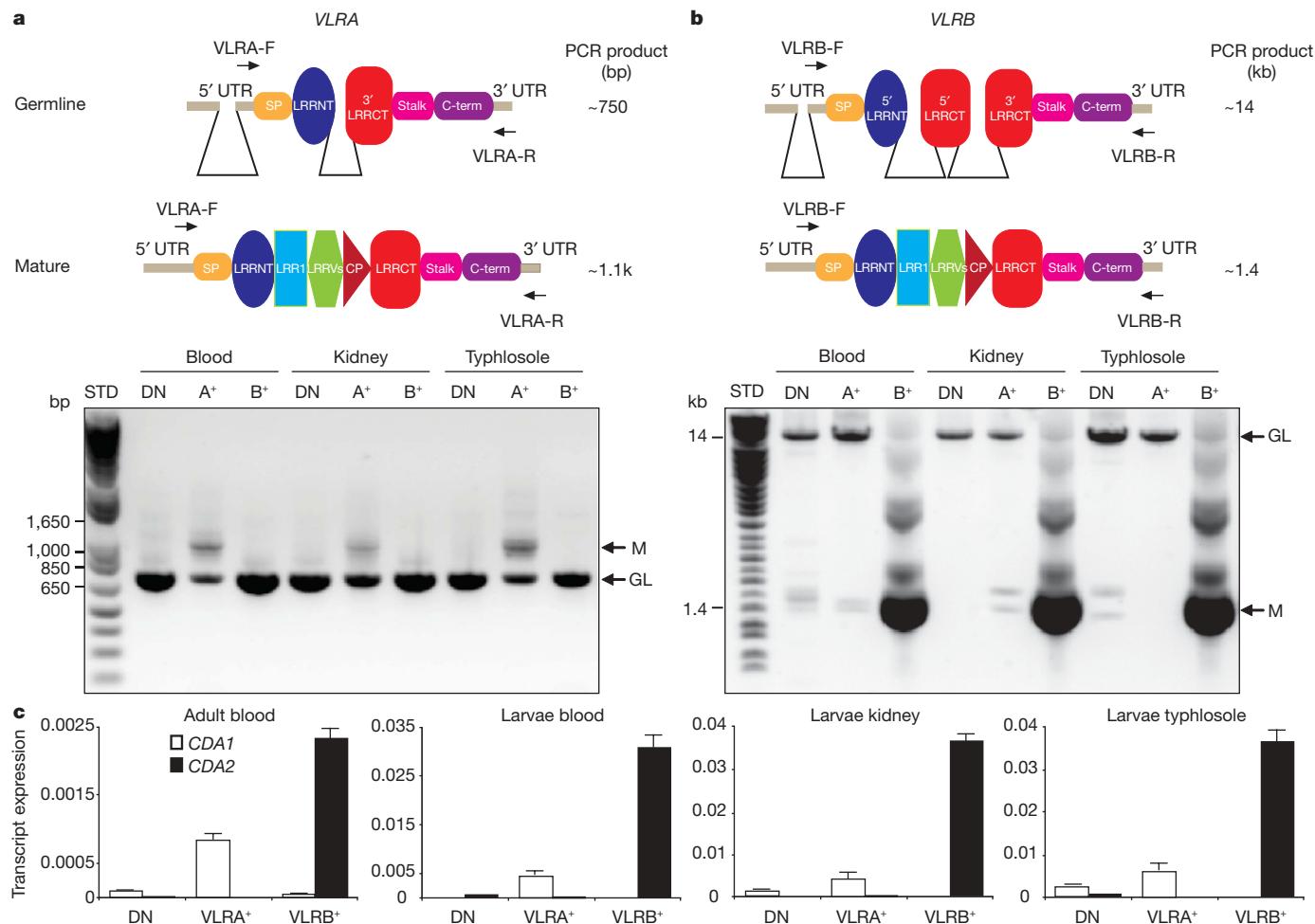


Figure 2 | Monoallelic assembly of VLRA and VLRB genes. **a, b**, Schematic of VLRA (**a**) and VLRB (**b**) genes before (top) and after (middle) gene assembly. Forward (F) and reverse (R) primer locations and the predicted sizes of PCR products are indicated. Lymphocytes from the indicated tissues were stained with anti-VLRA (R110) and anti-VLR-B (4C4) antibodies, and lymphocyte-gated cells were FACS sorted into three populations:

VLRA⁻ VLRB⁻ (DN), VLRA⁺ (A⁺) and VLRB⁺ (B⁺). VLRs were amplified from genomic DNA of the sorted lymphocyte populations (bottom). Germline (GL) and mature (M) products were verified by sequence analysis of representative DNA clones. bp, base pairs; kb, kilobases. **c**, CDA1 and CDA2 expression in sorted lymphocytes was measured by qPCR. Error bars indicate s.e.m., $n = 3$.

genomic sequence analysis, reveals gene expression profiles for VLRA and VLRB lymphocytes that are surprisingly similar to those for T and B cells in jawed vertebrates.

VLRA and VLRB lymphocyte responses to antigens

In response to immunization with *Bacillus anthracis* exosporium, antigen-binding VLRB⁺ members of the lymphocyte population undergo proliferation and differentiation into plasmacytes that secrete multivalent VLRB antibodies against BclA, a major protein in the outermost layer of *B. anthracis* spores²⁵. In the present experiments, we found that both the VLRA and VLRB lymphocyte populations responded to this antigen with comparable levels of proliferation (Fig. 4a). In contrast, VLRA lymphocytes that could bind anthrax spores were not demonstrable either before or after immunization (Fig. 4b); the same paradoxical results were obtained when *Escherichia coli* binding by VLRA and VLRB lymphocyte was examined before and after immunization (data not shown). Furthermore, the VLRA population failed to secrete VLRA proteins that could bind *B. anthracis*, *E. coli*, *Salmonella typhimurium* or *Streptococcus pneumoniae* after immunizations that led to the secretion of specific VLRB antibodies for each of these bacteria (Fig. 4c and Supplementary Fig. 6). Furthermore, VLRA proteins could not be detected in lamprey plasma by immunoblot analysis (Fig. 4d and Supplementary Fig. 7).

To determine whether an inherent difference in the two types of receptors could explain the failure to secrete VLRA, we transfected cells

of the 293T human embryonic kidney cell line with either VLRA or VLRB complementary DNAs to compare their patterns of intracellular synthesis and secretion. Although the VLRB-transfected cells secreted their VLRB protein products in the multimeric forms described previously²⁶, the VLRA-transfected cells expressed their VLRA protein products on the cell surface, but did not secrete them (Supplementary Fig. 8). These results indicate that VLRA lymphocytes respond to antigenic stimulation in a manner very reminiscent of the T-cell response, and confirm that VLRB lymphocytes respond to antigens much like B lymphocytes. In addition, the contrast between the abilities of VLRA and VLRB lymphocytes to bind native bacterial surface epitopes strongly suggests that they have different receptor repertoires.

Preferential PHA responsiveness of VLRA lymphocytes

Because of the resemblance between the VLRA lymphocytes and T lymphocytes in jawed vertebrates, we compared the responses of VLRA and VLRB lymphocytes to stimulation by the plant lectin phytohaemagglutinin (PHA), a classical T-cell mitogen. When larvae were injected with PHA, the VLRA⁺ lymphocytes responded vigorously by undergoing lymphoblastoid transformation and proliferation to become the predominant lymphocyte population (Fig. 5a, b). The activated VLRA⁺ cells were found by transmission electron microscopy to be large lymphoblasts with much less rough endoplasmic reticulum than the previously described VLRB-secreting plasmacytes (Fig. 5c)²⁵. Although VLRB lymphocytes also responded to PHA stimulation, their

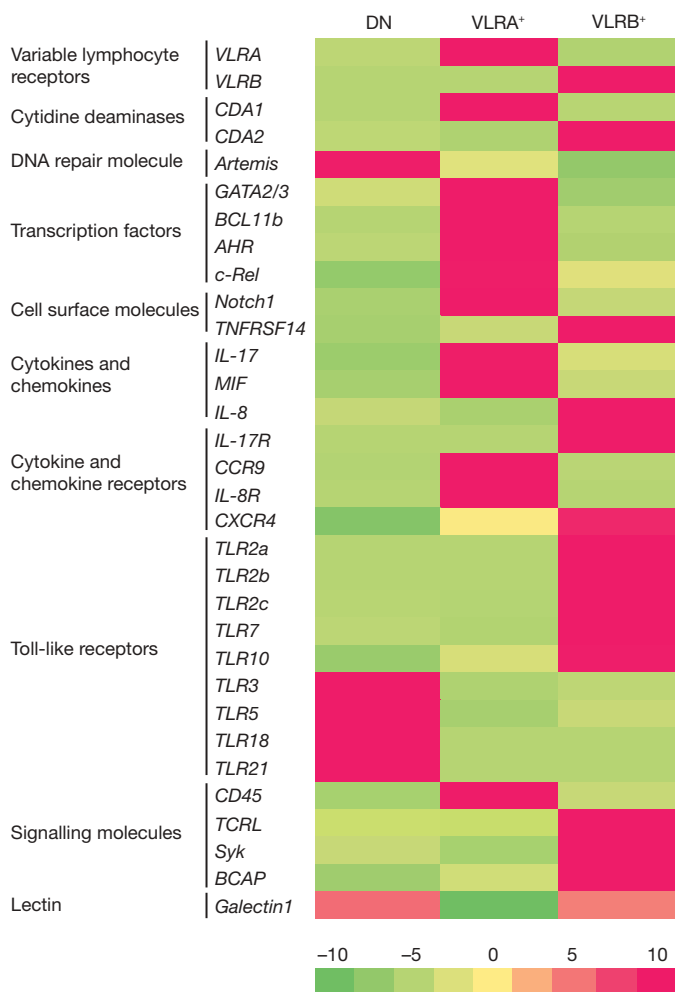


Figure 3 | Differential gene expression profiles of VLRA⁺ and VLRB⁺ lymphocytes. Lamprey lymphocytes were FACS sorted into three populations on the basis of VLRA and VLRB surface expression. Relative transcript levels of the indicated genes were measured by qPCR and compiled into a heat map as described in Methods.

response was much weaker in terms of cell size increases and the numbers of proliferating cells in this population. Nevertheless, PHA stimulation enhanced VLRB cell expression of *IL-8* transcripts, in contrast to its enhancement of VLRA cell expression of *IL-17* and *MIF* transcripts (Fig. 5d). These findings, coupled with the reciprocal expression of *IL-8R* by VLRA lymphocytes and *IL-17R* expression by VLRB lymphocytes (Fig. 3), are indicative of the potential for functional interactions between the two lymphocyte populations.

Discussion

This unforeseen genotypic and functional division of lymphocyte differentiation in lampreys offers an intriguing new piece to the puzzle of how the adaptive immune system may have evolved. Our initial definition of the VLRA and VLRB lymphocyte populations also offers interesting clues about key aspects of their differentiation and function, which are at present unresolved.

A question that is immediately raised by our findings concerns the anatomical location(s) in which VLRA lymphocytes begin their differentiation. Indeed, one of the reasons why the remarkable resemblance of VLRA lymphocytes to T lymphocytes was so surprising is the thus far fruitless quest to identify a thymus equivalent in lampreys^{16,17}. However, the relatively high concentration of VLRA lymphocytes in the gill region could favour the branchial region as a potential site for their generation. Furthermore, our gene profile analysis indicates that VLRA lymphocytes express an array of transcription factors,

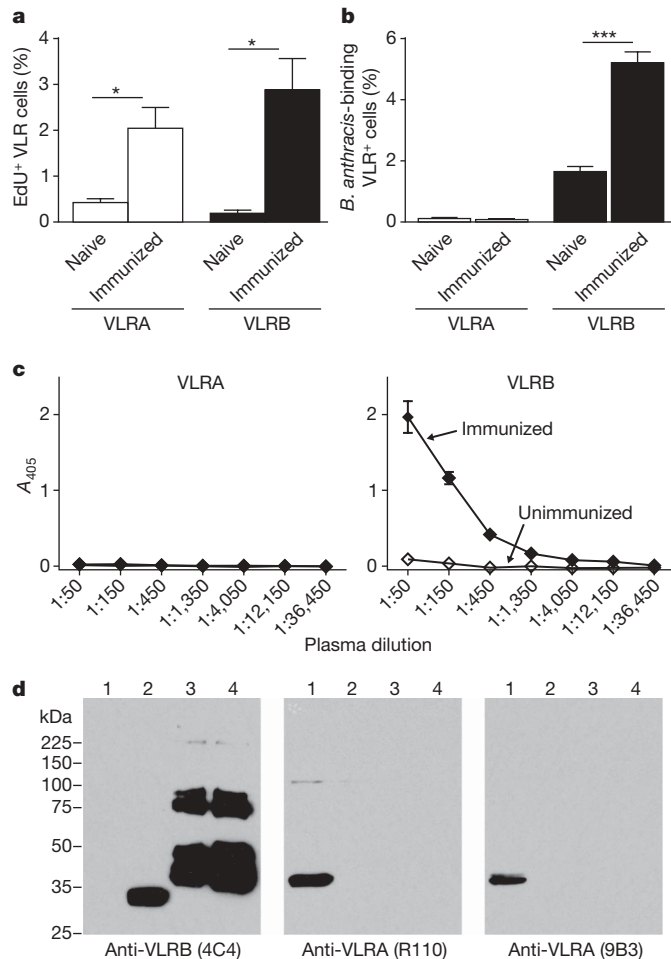


Figure 4 | Antigen-activated VLRA⁺ lymphocytes do not secrete their receptors. **a**, Immunization with *B. anthracis* exosporium induces proliferation of VLRA⁺ and VLRB⁺ lymphocytes; $n = 5$; $*P < 0.05$. EdU, 5-ethynyl-2'-deoxyuridine. **b**, VLRA⁺ lymphocytes do not bind to *B. anthracis* spores before immunization or 14 days after booster immunization; $n = 5$; $***P < 0.001$. **c**, **d**, VLRB, but not VLRA, is secreted. **c**, *Bacillus-anthraxis*-specific VLRA and VLRB reactivity was evaluated by ELISA with anti-VLRA (9A6) and anti-VLRB (4C4) monoclonal antibodies. **d**, Immunoblot of VLRA (lane 1) and VLRB (lane 2) HEK-293T cell transfectant lysates, naive lamprey plasma (lane 3), and immunized plasma (lane 4) under reducing conditions. 9B3 denotes anti-VLRA monoclonal antibody; R110 denotes anti-VLRA rabbit antiserum. Error bars indicate s.e.m.

chemokine receptors and Notch1 that lymphocyte progenitors use for homing to the thymus and T lineage commitment in jawed vertebrates. Analysis of the expression of the corresponding chemokines and Notch1 ligands in the pharyngeal pouch-derived epithelial cells and the neighbouring VLRA lymphocytes could therefore facilitate the identification of a thymus equivalent in lampreys.

Solving the issue of where the VLRA and VLRB genes are assembled will also require better understanding of the molecular mechanisms involved in this process. The donor and recipient LRR sequences used in VLR assembly lack recombination signal sequences that characterize the immunoglobulin variable, diversity and joining gene segments, and the RAG1 and RAG2 recombination activation genes used for TCR and BCR gene assembly have not been found in lampreys¹⁸. The features of VLR gene assembly known at present instead suggest that a gene conversion mechanism is used to copy the donor LRR sequences^{21,23,24}. In accord with this hypothesis, two AID/APOBEC family orthologues, CDA1 and CDA2, have been identified in the lamprey. CDA1 has been shown to have an AID-like DNA deaminase mutator function in bacteria and yeast lacking uracil DNA glycosylase activity²¹. Although the functional competency of CDA2 has not yet been elaborated, it shares the DNA deaminase catalytic motif with

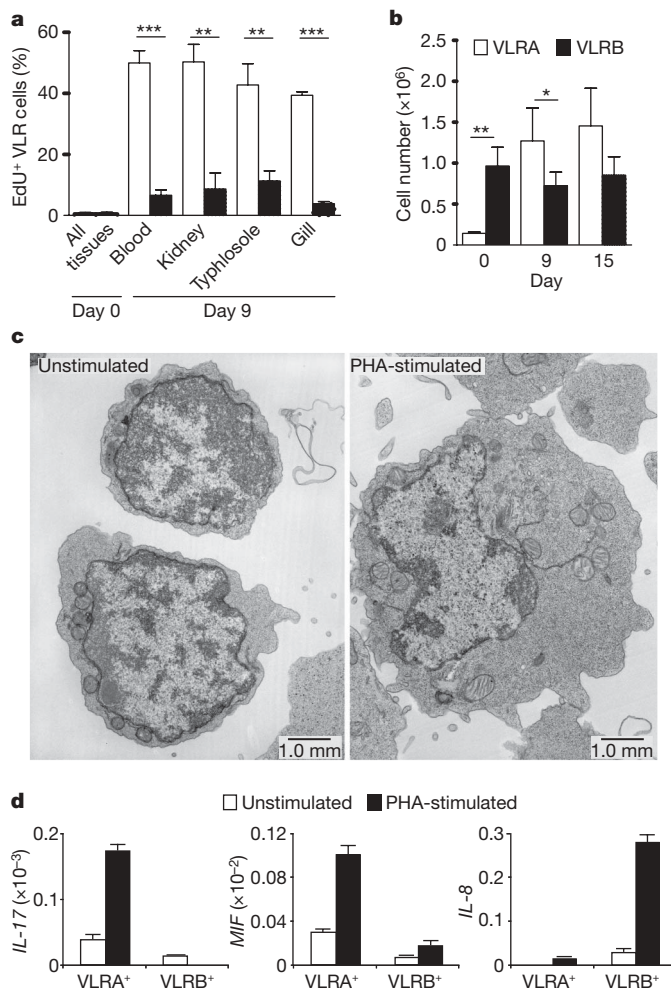


Figure 5 | PHA preferentially stimulates VLRA⁺ lymphocytes. **a, b,** *In vivo* PHA stimulation. Lampreys were given 25 μg of PHA by intracoelomic injection on day 0. Proliferation (**a**) and absolute numbers (**b**) of VLRA⁺ and VLRB⁺ cells were determined on days 9 and 15 after injection; *n* = 4 (**a**) and 7 (**b**); **P* < 0.05, ***P* < 0.01 and ****P* < 0.001. **c,** Transmission electron microscopy imaging of unstimulated (left panel) and PHA-stimulated (day 9) (right panel) VLRA⁺ cells. PHA-stimulated cells have a larger diameter (mean ± s.e.m., naive = 3.8 ± 0.1 μm (*n* = 7) versus PHA-stimulated = 7.4 ± 0.3 μm (*n* = 7, *P* < 0.01)) and increased cytoplasm. **d,** Cytokine expression by VLRA⁺ and VLRB⁺ lymphocytes. FACS-sorted blood lymphocytes were isolated from PHA-stimulated (day 7, *n* = 3) or unstimulated lampreys (*n* = 3), and analysed for expression of *IL-17* (left panel), *MIF* (middle panel), and *IL-8* (right panel) by qPCR. Error bars indicate s.e.m.

CDA1 and also has an extended C terminus containing a small AT-hook peptide motif that may confer DNA binding specificity. These distinctive features, and the restriction of CDA1 and CDA2 expression to VLRA and VLRB lymphocytes, respectively, thus imply the potential for selective involvement of these enzymes in VLRA and VLRB assembly during lymphocyte development.

One of the most provocative issues raised by our findings concerns the mode of antigen recognition by VLRA-bearing lymphocytes. We expected to find comparable antigen binding and response patterns for the two lymphocyte populations, because the VLRA and VLRB proteins are very similar and have comparable diversity. The proliferative responses of VLRA and VLRB lymphocytes to immunization indeed were found to be equivalent. Surprisingly however, antigen-binding VLRA lymphocytes could not be found before or after immunization, whereas specific antigen-binding VLRB lymphocytes were easily detected in naive animals, and immunization increased their frequency. A possible explanation for the inability of VLRA lymphocytes to bind native epitopes on immunogens that induce

their proliferation is that VLRA lymphocytes recognize processed antigens and undergo selection in a manner analogous to the T-lymphocyte repertoire selection in jawed vertebrates. Lampreys lack the MHC genes for MHC class I and class II molecules that accessory cells use to present peptide fragments to T cells in jawed vertebrates¹⁸. However, a precedent for an alternative mechanism for self recognition is provided by the recent identification of a highly polymorphic fusion/histocompatibility gene locus in *Botryllus schlosseri*⁴⁹, a colonial tunicate member of a vertebrate sister group in the new chordate phylogeny⁵⁰. The nature of the antigens recognized by VLRA anticipatory receptors remains an unresolved enigma at present.

The failure of VLRA cells to secrete their receptors as soluble proteins was the most notable functional difference observed between the VLRA and VLRB lymphocytes. This feature could reflect an inability of VLRA cells to develop the necessary secretory machinery. However, when VLRA proteins were produced by a human cell line that can secrete VLRB proteins, the transfected cells produced cell surface VLRA proteins, but did not secrete them. The VLRA proteins, like the TCR proteins, thus seem to be expressed exclusively as transmembrane molecules. The preferential expression of *TLR2*, *TLR7* and *TLR10* orthologues by the VLRB lymphocytes suggests that TLR ligands may facilitate activation of this population of lymphocytes in a manner similar to their roles in activating B lymphocytes. The development of appropriate conditions for culturing lamprey lymphocytes and cell lines will facilitate the future analysis of the signalling mechanisms used by VLRA and VLRB lymphocytes.

Many of the elaborate cytokine networks that T and B cells use to coordinate effective inflammatory responses in jawed vertebrates are not evident in the more basal jawless vertebrates. Agnathan orthologues have not been identified for the interferons and interferon receptors *IL-2/IL-2R*, *IL-4/IL4R*, *IL-7/IL-7R* and other cytokine/cytokine receptor pairs. Nevertheless, the VLRA lymphocytes preferentially express *IL-17* and *MIF* transcripts, and their enhanced expression after activation suggests that VLRA lymphocytes may use these pro-inflammatory cytokines to engage the *IL-17R*-bearing VLRB lymphocytes for defence purposes, in a manner similar to their T cell counterparts. Conversely, activated VLRB lymphocytes upregulate their expression of *IL-8* transcripts, suggesting that they may use this cytokine to attract and engage *IL-8R*-bearing VLRA lymphocytes and other cell types to amplify the response to potential pathogens.

The evolution around 500 million years ago of two very different anticipatory receptors of comparable diversity in jawless and jawed vertebrates, while conserving similar compartmentalization of lymphocyte differentiation, strongly attests the survival value of adaptive immunity. The evolution of an LRR-based adaptive immune system in lampreys and hagfish, apparently before the second round of whole genomic replication, and an immunoglobulin-based adaptive immune system in jawed vertebrates, after a second round of genomic duplication⁵⁰, suggests that these very different types of anticipatory receptors were convergent evolutionary acquisitions¹⁹. The failure so far to find relics of the VLR genes in jawed vertebrates is also consistent with an independent evolution of clonally diverse LRR-based receptors versus immunoglobulin-based receptors for otherwise similar adaptive immune systems. An obvious advantage of adaptive immunity over the elaborate invertebrate strategies for innate immunity is the capacity for memory of previous encounters with infectious agents. On the other hand, the development of a randomly generated anticipatory repertoire of great diversity inevitably creates the hazard of self-reactivity and attendant autoimmunity. The orchestration of cellular and humoral immune responses by T cells, whose TCR repertoire is shaped during development to discriminate between self and non-self antigens, provides an important safeguard against autoimmunity in jawed vertebrates. The potential for autoimmunity theoretically would necessitate a similar solution for the adaptive immune system in jawless vertebrates. Our findings thus suggest that dual recognition and response arms with intertwined function were fundamental to the evolution of adaptive immunity in vertebrates.

METHODS SUMMARY

Sea lamprey larvae (8–15 cm, 2–4 years of age) were collected from tributaries to Lake Michigan (Lamprey services) and housed in sand-lined aquariums at 20 °C. Larvae anaesthetized with 0.1 g l⁻¹ MS222 (Sigma) were given intracoelomic injections of antigens or mitogens (25 µg phytohaemagglutinin (PHA)-L, Sigma) prepared in 60 µl of sterile 0.67× PBS. Injections were administered on days 0 and 14, and tissues were collected from lampreys euthanized with 1 g l⁻¹ MS222 on day 28 (unless otherwise indicated). Blood was collected in 30 mM EDTA, 0.67× PBS. Buffy coat leukocytes were separated from red blood cells by centrifugation at 50g. Leukocytes were isolated from kidney, typhlosole and gills by disrupting tissues between frosted glass slides.

For VLRA antibody production, four VLRA cDNAs isolated from lamprey lymphocytes were cloned in-frame with the constant region of human IgG1 (IgG1-Fc). The VLRA-IgG1-Fc fusion proteins were expressed in HEK-293T cells and purified from tissue culture supernatants by protein A (GE Healthcare) chromatography. Monoclonal anti-VLRA antibodies were produced by immunization of BALB/c mice with VLRA-IgG1-Fc proteins emulsified in TiterMax Gold adjuvant. Lymphocytes from the draining lymph nodes of immunized mice were fused with the Ag8.653 myeloma cell line using PEG-1500 (Roche). Three VLRA-specific hybridoma clones (9A6 (IgG1), 9B3 (IgG1) and 2A5 (IgG1)) were identified by ELISA with VLRA recombinant protein and immunoblotting of VLRA transfectants. Anti-VLRA polyclonal antisera were produced by immunization of rabbits with the VLRA-IgG1-Fc proteins (PickCell Laboratories BV).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 February; accepted 14 April 2009.

Published online 27 May 2009.

- Paul, W. E. *Fundamental Immunology* 6th edn (Lippincott Williams & Wilkins, 2008).
- Cooper, M. D., Peterson, R. D. & Good, R. A. Delineation of the thymic and bursal lymphoid systems in the chicken. *Nature* **205**, 143–146 (1965).
- Roitt, I. M., Greaves, M. F., Torigiani, G., Brostoff, J. & Playfair, J. H. The cellular basis of immunological responses. A synthesis of some current views. *Lancet* **2**, 367–371 (1969).
- Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575–581 (1983).
- Yanagi, Y. et al. A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature* **308**, 145–149 (1984).
- Hedrick, S. M., Cohen, D. I., Nielsen, E. A. & Davis, M. M. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* **308**, 149–153 (1984).
- Jung, D., Giallourakis, C., Mostoslavsky, R. & Alt, F. W. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu. Rev. Immunol.* **24**, 541–570 (2006).
- Burnet, F. M. A modification of Jerne's theory of antibody production using the concept of clonal selection. *Aust. J. Sci.* **20**, 67–69 (1957).
- Zinkernagel, R. M. & Doherty, P. C. Cytotoxic thymus-derived lymphocytes in cerebrospinal fluid of mice with lymphocytic choriomeningitis. *J. Exp. Med.* **138**, 1266–1269 (1973).
- Nikolic-Zugic, J. & Bevan, M. J. Role of self-peptides in positively selecting the T-cell repertoire. *Nature* **344**, 65–67 (1990).
- von Boehmer, H. Selection of the T-cell repertoire: receptor-controlled checkpoints in T-cell development. *Adv. Immunol.* **84**, 201–238 (2004).
- Unanue, E. R. Perspective on antigen processing and presentation. *Immunol. Rev.* **185**, 86–102 (2002).
- Flajnik, M. F. Comparative analyses of immunoglobulin genes: surprises and portents. *Nature Rev. Immunol.* **2**, 688–698 (2002).
- Litman, G. W., Cannon, J. P. & Dishaw, L. J. Reconstructing immune phylogeny: new perspectives. *Nature Rev. Immunol.* **5**, 866–879 (2005).
- Finstad, J. & Good, R. A. The evolution of the immune response: III. immunologic responses in the lamprey. *J. Exp. Med.* **120**, 1151–1168 (1964).
- Ardavin, C. F. & Zapata, A. The pharyngeal lymphoid tissue of lampreys. A morpho-functional equivalent of the vertebrate thymus? *Thymus* **11**, 59–65 (1988).
- Boehm, T. & Bleul, C. C. The evolutionary history of lymphoid organs. *Nature Immunol.* **8**, 131–135 (2007).
- Unik-Ool, T. et al. Lamprey lymphocyte-like cells express homologs of genes involved in immunologically relevant activities of mammalian lymphocytes. *Proc. Natl Acad. Sci. USA* **99**, 14356–14361 (2002).
- Pancer, Z. et al. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* **430**, 174–180 (2004).
- Pancer, Z. et al. Variable lymphocyte receptors in hagfish. *Proc. Natl Acad. Sci. USA* **102**, 9224–9229 (2005).
- Rogozin, I. B. et al. Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nature Immunol.* **8**, 647–656 (2007).
- Alder, M. N. et al. Diversity and function of adaptive immune receptors in a jawless vertebrate. *Science* **310**, 1970–1973 (2005).
- Nagawa, F. et al. Antigen-receptor genes of the agnathan lamprey are assembled by a process involving copy choice. *Nature Immunol.* **8**, 206–213 (2007).
- Cooper, M. D. & Alder, M. N. The evolution of adaptive immune systems. *Cell* **124**, 815–822 (2006).
- Alder, M. N. et al. Antibody responses of variable lymphocyte receptors in the lamprey. *Nature Immunol.* **9**, 319–327 (2008).
- Herrin, B. R. et al. Structure and specificity of lamprey monoclonal antibodies. *Proc. Natl Acad. Sci. USA* **105**, 2040–2045 (2008).
- Ho, I. C., Tai, T. S. & Pai, S. Y. GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. *Nature Rev. Immunol.* **9**, 125–135 (2009).
- Suzuki, T., Shin, I. T., Kohara, Y. & Kasahara, M. Transcriptome analysis of hagfish leukocytes: a framework for understanding the immune system of jawless fishes. *Dev. Comp. Immunol.* **28**, 993–1003 (2004).
- Banerjee, D., Liou, H. C. & Sen, R. c-Rel-dependent priming of naive T cells by inflammatory cytokines. *Immunity* **23**, 445–458 (2005).
- Quintana, F. J. et al. Control of T(reg) and T(H)17 cell differentiation by the aryl hydrocarbon receptor. *Nature* **453**, 65–71 (2008).
- Wakabayashi, Y. et al. Bcl11b is required for differentiation and survival of alphabeta T lymphocytes. *Nature Immunol.* **4**, 533–539 (2003).
- Liu, C. et al. Coordination between CCR7- and CCR9-mediated chemokine signals in prevascular fetal thymus colonization. *Blood* **108**, 2531–2539 (2006).
- Maillard, I., Fang, T. & Pear, W. S. Regulation of lymphoid development, differentiation, and function by the Notch pathway. *Annu. Rev. Immunol.* **23**, 945–974 (2005).
- Kishihara, K. et al. Normal B lymphocyte development but impaired T cell maturation in CD45-exon6 protein tyrosine phosphatase-deficient mice. *Cell* **74**, 143–156 (1993).
- Lippert, U., Zachmann, K., Henz, B. M. & Neumann, C. Human T lymphocytes and mast cells differentially express and regulate extra- and intracellular CXCR1 and CXCR2. *Exp. Dermatol.* **13**, 520–525 (2004).
- Weaver, C. T., Hatton, R. D., Mangan, P. R. & Harrington, L. E. IL-17 family cytokines and the expanding diversity of effector T cell lineages. *Annu. Rev. Immunol.* **25**, 821–852 (2007).
- Tsutsui, S., Nakamura, O. & Watanabe, T. Lamprey (*Lethenteron japonicum*) IL-17 upregulated by LPS-stimulation in the skin cells. *Immunogenetics* **59**, 873–882 (2007).
- Bloom, B. R. & Shevach, E. Requirement for T cells in the production of migration inhibitory factor. *J. Exp. Med.* **142**, 1306–1311 (1975).
- Weiser, W. Y. et al. Molecular cloning of a cDNA encoding a human macrophage migration inhibitory factor. *Proc. Natl Acad. Sci. USA* **86**, 7522–7526 (1989).
- Ma, Q., Jones, D. & Springer, T. A. The chemokine receptor CXCR4 is required for the retention of B lineage and granulocytic precursors within the bone marrow microenvironment. *Immunity* **10**, 463–471 (1999).
- Murphy, K. M., Nelson, C. A. & Sedy, J. R. Balancing co-stimulation and inhibition with BTLA and HVEM. *Nature Rev. Immunol.* **6**, 671–681 (2006).
- Jumaa, H., Hendriks, R. W. & Reth, M. B cell signaling and tumorigenesis. *Annu. Rev. Immunol.* **23**, 415–445 (2005).
- Okada, T., Maeda, A., Iwamatsu, A., Gotoh, K. & Kurosaki, T. BCAP: the tyrosine kinase substrate that connects B cell receptor to phosphoinositide 3-kinase activation. *Immunity* **13**, 817–827 (2000).
- Yu, C., Ehrhardt, G. R., Alder, M. N., Cooper, M. D. & Xu, A. Inhibitory signaling potential of a TCR-like molecule in lamprey. *Eur. J. Immunol.* **39**, 571–579 (2009).
- Sims-Mourtada, J. C. et al. *In vivo* expression of interleukin-8, and regulated on activation, normal, T-cell expressed, and secreted, by human germinal centre B lymphocytes. *Immunology* **110**, 296–303 (2003).
- Pasare, C. & Medzhitov, R. Control of B-cell responses by Toll-like receptors. *Nature* **438**, 364–368 (2005).
- Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124**, 783–801 (2006).
- Ishii, A. et al. Lamprey TLRs with properties distinct from those of the variable lymphocyte receptors. *J. Immunol.* **178**, 397–406 (2007).
- De Tomaso, A. W. et al. Isolation and characterization of a protochordate histocompatibility locus. *Nature* **438**, 454–459 (2005).
- Putnam, N. H. et al. The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064–1071 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. N. Alder and G. R. A. Ehrhardt for suggestions and discussion; C. L. Turnbough, Jr for providing *B. anthracis* spores and exosporium; D. E. Briles and W. H. Benjamin, Jr for *E. coli*, *S. pneumoniae* and *S. typhimurium*; H. Yi for help with electron microscopy; S. A. Durham and R. E. Karaffa, II for help with cell sorting; M. Flurry for help with preparation of figures. This work is supported by the National Institutes of Health and the Georgia Research Alliance.

Author Contributions P.G., M.H., B.R.H., J.L., C.Y., A.S. and M.D.C. designed the research, analysed data and wrote the paper; P.G., M.H., B.R.H., J.L., C.Y. and A.S. performed the research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.D.C. (max.cooper@emory.edu).

METHODS

Antigens. *Escherichia coli* (ATCC 25922 strain), *Salmonella enterica* Typhimurium (WB99-SR11 strain) and *Streptococcus pneumoniae* (R36A strain) were provided by D. E. Briles and W. H. Benjamin, Jr (University of Alabama at Birmingham). Bacteria were heat-killed by incubation at 60 °C for 1 h. Five-hundred-million heat-killed *E. coli* and *S. typhimurium*, and 1×10^8 *S. pneumoniae* were used for immunization. *Bacillus anthracis* (spores and exosporium) and *Bacillus cereus* spores were provided by C. L. Turnbough, Jr (University of Alabama at Birmingham). Lampreys were immunized with 10 µg of *B. anthracis* exosporium.

Flow cytometric analysis of VLRA and VLRB lymphocytes. Leukocytes from blood and tissues were stained with anti-VLRA rabbit polyclonal serum (R110 or R111) or mouse monoclonal antibodies (9A6, 9B3, 2A5, IgG1) and anti-VLRB mouse monoclonal antibody (4C4, IgG2b) for 10 min on ice, washed twice and stained with fluorescent-conjugated secondary antibodies. For anti-VLRA (IgG1) and anti-VLRB (4C4, IgG2b) antibody staining combinations, R-phycoerythrin (PE)-conjugated goat anti-mouse IgG1 antibodies (Southern Biotech) and FITC-conjugated goat anti-mouse IgG2b antibodies (Southern Biotech) were used. When staining with anti-VLRA rabbit serum and anti-VLRB monoclonal antibody (4C4), RPE-conjugated goat anti-rabbit Ig antibodies (Southern Biotech) and allophycocyanin (APC)-conjugated goat anti-mouse Ig antibodies (Southern Biotech) were used. Cells were analysed on a CyAn ADP high-performance flow cytometer (Dako). Staining and washes were in $0.67 \times$ PBS with 1% BSA.

Immunoblotting. Samples were separated on 10% SDS-PAGE gels with or without 5% (v/v) 2-mercaptoethanol before transfer onto nitrocellulose membranes. Membranes were blocked overnight with 5% milk and incubated with anti-VLRA or anti-VLRB antibodies for 1 h. After five washes with PBS-0.5% Tween-20, membranes were incubated with HRP-conjugated goat anti-rabbit or mouse Ig polyclonal antibodies (Southern Biotech) and washed. Blots were developed using SuperSignal West Pico chemiluminescent substrate (Pierce).

Genomic DNA and PCR. Genomic DNA was extracted from VLRA⁺, VLRB⁺ and VLR double-negative-sorted cells in blood and tissue lymphocytes using DNeasy kit (QIAGEN). Genomic PCR was carried out using primers VLRA-F and VLRA-R (Expand High Fidelity, Roche, and Ex Taq, Takara) or VLRB-F and VLRB-R (Expand Long Template, Roche). Primers are listed in Supplementary Table 1.

Quantitative PCR. Target gene sequences were obtained from the National Center for Biotechnology Information database or the lamprey genome database of the Genome Sequencing Center at Washington University. Cells in the lymphocyte gate were sorted into VLRA⁺, VLRB⁺ and VLR double-negative populations on a BD FACS Aria II. RNA was extracted from each population using RNeasy kit with on-column DNA digestion by DNase I (QIAGEN). First-strand cDNA was synthesized with random hexamer primers and Superscript III (Invitrogen). Quantitative real-time PCR was done with SYBR Green on a 7900HT ABI Prism (Applied Biosystems). Three independent quantitative real-time PCR reactions were performed at different times for each sample. The value of the target gene was normalized to β-actin. The normalized value of each gene was compiled into a heat map (z -score $\times 10$). Magenta: $z > 0$, green:

$z < 0$ and yellow: $z = 0$ [$z = (\text{each value} - \text{average})/\text{s.d.}$]. Primers are described in Supplementary Table 1.

Proliferation assay. Lampreys stimulated with antigen or mitogen were injected with 5 µg of 5-ethynyl-2'-deoxyuridine (Invitrogen) in 60 µl $0.67 \times$ PBS and returned to their tanks for 24 h before collection of leukocytes for staining with anti-VLRA or anti-VLRB antibodies. Stained cells were fixed in 4% para-formaldehyde for 15 min at room temperature, then resuspended in 100 µl of the $1 \times$ saponin-based permeabilization and wash buffer for 10 min on ice. EdU⁺ cell detection was performed according to the manufacturer's protocol (Click-iT EdU Alexa Fluor 488 Flow Cytometry Assay Kit, Invitrogen). In brief, cells were incubated with the EdU reaction cocktail, which contains Alexa Fluor 488-azide, for 30 min at room temperature, washed twice in $1 \times$ saponin-based permeabilization and wash buffer, and analysed by flow cytometry.

ELISA. For animals immunized with *B. anthracis* exosporium, 10^6 *B. anthracis* spores or *B. cereus* strain T spores per well were applied on poly-L-lysine-coated plates as described previously²². For animals immunized with bacteria, 5×10^6 heat-killed bacteria per well were applied on poly-L-lysine-coated plates. Plates were blocked with 1% BSA in PBS for 1 h. Serially diluted plasma was applied for 2 h at 37 °C. VLRA were detected with anti-VLRB (4C4) or anti-VLRA (9A6) monoclonal antibodies and alkaline phosphatase-conjugated goat anti-mouse Ig antibodies (Southern Biotech). ELISAs were developed with phosphatase substrate (Sigma) and read at 405 nm (Versamax microplate reader, Molecular Devices).

Lymphocyte antigen-binding assay. One million lamprey leukocytes were incubated with 10^6 Alexa-Fluor-488-labelled spores or 10^7 Alexa-Fluor-488-labelled bacteria for 20 min on ice. Spores and bacteria were labelled using the Alexa Fluor 488 Protein Labelling Kit (Invitrogen). Cells were stained with anti-VLRA (R110) and anti-VLRB (4C4) antibodies followed by APC-conjugated goat anti-mouse Ig antibodies (Southern Biotech) and RPE-conjugated goat anti-rabbit Ig antibodies (Southern Biotech), then analysed by flow cytometry.

VLRA expression in HEK-293T cells. VLRA and VLRB plasmids were transfected into HEK-293T cells cultured in DMEM containing 5% FBS using linear poly-ethylenimine (PEI), MW 25,000 (Polysciences, Inc.) at a 3:1 PEI:DNA ratio. Cells were separated from supernatants 48 h after transfection by centrifugation at 300g and lysed in 1% NP-40 lysis buffer.

Lymphocyte counts. Lymphocytes were extracted from lamprey kidney and typhlosol by passage through a 70-µm cell strainer (BD Bioscience) to obtain a single cell suspension. Cells were washed and resuspended in 1 ml $0.67 \times$ PBS. Total cells were counted in 25 µl of each sample on an Accuri C6 flow cytometer (Accuri Cytometers Inc.). The total number of cells in the 'lymphocyte gate' was calculated by the formula: total lymphocyte number = number of cells in the lymphocytes gate $\times 40$ (dilution factor: 1,000 µl/25 µl). Lymphocytes were then stained with anti-VLRA and -VLRB antibodies to determine the percentage of VLRA⁺ and VLRB⁺ cells in the samples. These percentages were used to calculate the total number of lymphocytes of each type.

Electron microscopy. Blood lymphocytes from naive or PHA-stimulated lamprey were sorted on BD FACS Aria II (BD Bioscience). The VLRA-positive cells were prepared for transmission electron microscopic analysis as described²².

Statistical analysis. A two-sample Student's *t*-test was used for statistical analysis.

ARTICLES

Histone H4 lysine 16 acetylation regulates cellular lifespan

Weiwei Dang¹, Kristan K. Steffen², Rocco Perry¹, Jean A. Dorsey¹, F. Brad Johnson⁴, Ali Shilatifard⁵, Matt Kaeberlein³, Brian K. Kennedy² & Shelley L. Berger^{1,6}

Cells undergoing developmental processes are characterized by persistent non-genetic alterations in chromatin, termed epigenetic changes, represented by distinct patterns of DNA methylation and histone post-translational modifications. Sirtuins, a group of conserved NAD⁺-dependent deacetylases or ADP-ribosyltransferases, promote longevity in diverse organisms; however, their molecular mechanisms in ageing regulation remain poorly understood. Yeast Sir2, the first member of the family to be found, establishes and maintains chromatin silencing by removing histone H4 lysine 16 acetylation and bringing in other silencing proteins. Here we report an age-associated decrease in Sir2 protein abundance accompanied by an increase in H4 lysine 16 acetylation and loss of histones at specific subtelomeric regions in replicatively old yeast cells, which results in compromised transcriptional silencing at these loci. Antagonizing activities of Sir2 and Sas2, a histone acetyltransferase, regulate the replicative lifespan through histone H4 lysine 16 at subtelomeric regions. This pathway, distinct from existing ageing models for yeast, may represent an evolutionarily conserved function of sirtuins in regulation of replicative ageing by maintenance of intact telomeric chromatin.

Cells undergo characteristic molecular alterations as organisms age^{1,2}. Studies in model organisms identify conserved genetic pathways that modulate ageing, such as insulin signalling, oxidative stress tolerance and nutrient sensing^{3,4}. Ageing and other developmental processes, such as differentiation, apoptosis and gametogenesis, associate with characteristic epigenetic changes at the cellular level, including DNA methylation and histone post-translational modifications^{5–7}. Nevertheless, the functions of these molecular changes during ageing remain unknown. Although sirtuins promote longevity in yeast, worms and flies⁸, conserved pathways for sirtuins in ageing regulation remain controversial. Yeast *Saccharomyces cerevisiae* silencing information regulator 2 (Sir2), the first sirtuin to be found, establishes and maintains silencing within yeast heterochromatic-like regions at telomeres, ribosomal DNA (rDNA) and silenced mating-type loci by removing H4 lysine 16 (H4K16) acetylation and bringing in other silencing proteins⁹. Antagonizing activities of Sir2 and a histone acetyltransferase, Sas2, generate a gradient of H4K16 acetylation marking the boundary of silencing chromatin near telomeres^{10,11}.

One cause of yeast ageing is nucleolar accumulation of extrachromosomal rDNA circles (ERCs), generated from recombination between rDNA repeats as cells divide¹². Deletion of *SIR2* results in hyper-recombination within the rDNA and elevated levels of ERCs¹³, whereas *FOB1* deletion reduces rDNA recombination and ERC formation and extends lifespan^{14,15}. Overexpression of Sir2 also increases lifespan, but does not further increase the lifespan when combined with deletion of *FOB1*, which has been interpreted to suggest that Sir2 overexpression promotes longevity, in part, by repressing ERCs¹³. Sir2 overexpression also increases lifespan in worms and flies⁸; however, ERCs have not been detected in these organisms or in mammalian cells in association with ageing, and their accumulation is thus unlikely to underlie ageing in diverse organisms^{16,17}.

A central question is whether histone acetylation levels are altered during ageing and whether such changes contribute directly to ageing. Although histones are almost the only acetylated proteins reported in yeast¹⁸, there are numerous acetylated proteins in metazoans¹⁹; in addition, deacetylases, including the sirtuin class, target both histones and non-histone proteins²⁰. In mammalian cells, SIRT1, SIRT2, SIRT3 and SIRT6 deacetylate histones^{21–23} but have non-histone substrates, and some of these are promising candidates in ageing pathways. For example, p53, which promotes apoptosis, is deacetylated by SIRT1, extending cell survival²³. Hence, despite extensive knowledge about the sirtuin class of deacetylases, it remains to be established whether deacetylation of H4K16 plays an important role in the Sir2 ageing pathway.

In this study, we investigated age-associated changes in histone modifications and chromatin. In particular, because ERC accumulation in the nucleolus is not conserved through evolution, we explored chromatin-related changes at Sir2-regulated loci outside of the nucleoli. Our results establish chromatin as a critical target for Sir2 in regulation of lifespan and indicate that Sir2 opposes replicative ageing in yeast through histone targets located near telomeres.

Histone modifications change in old cells

We examined histone post-translational modifications in young and old cells, by means of isolation of biotin-labelled mother cells²⁴. Progressive ageing was evident with increasing bud scars and rDNA copy numbers¹² (Supplementary Fig. 1a, b). Histone modifications associated with gene transcription, DNA repair and chromosome condensation were tested by western blotting. Notably, the level of H4K16 acetylation increased with age (Fig. 1a, left panel), whereas H3K56 acetylation decreased (Fig. 1a, middle panel). No significant changes were found for other modifications and histone variants tested (Fig. 1a and Supplementary Fig. 2). Sir2 deacetylates

¹Gene Expression and Regulation Program, The Wistar Institute Philadelphia, Pennsylvania 19104, USA. ²Department of Biochemistry, ³Department of Pathology, University of Washington Seattle, Washington 98195, USA. ⁴Department of Pathology and Laboratory Medicine, Cell and Molecular Biology Group, Biomedical Graduate Studies and Institute on Aging, University of Pennsylvania School of Medicine, Philadelphia 19104, Pennsylvania, USA. ⁵Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA. ⁶Department of Cell & Developmental Biology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.

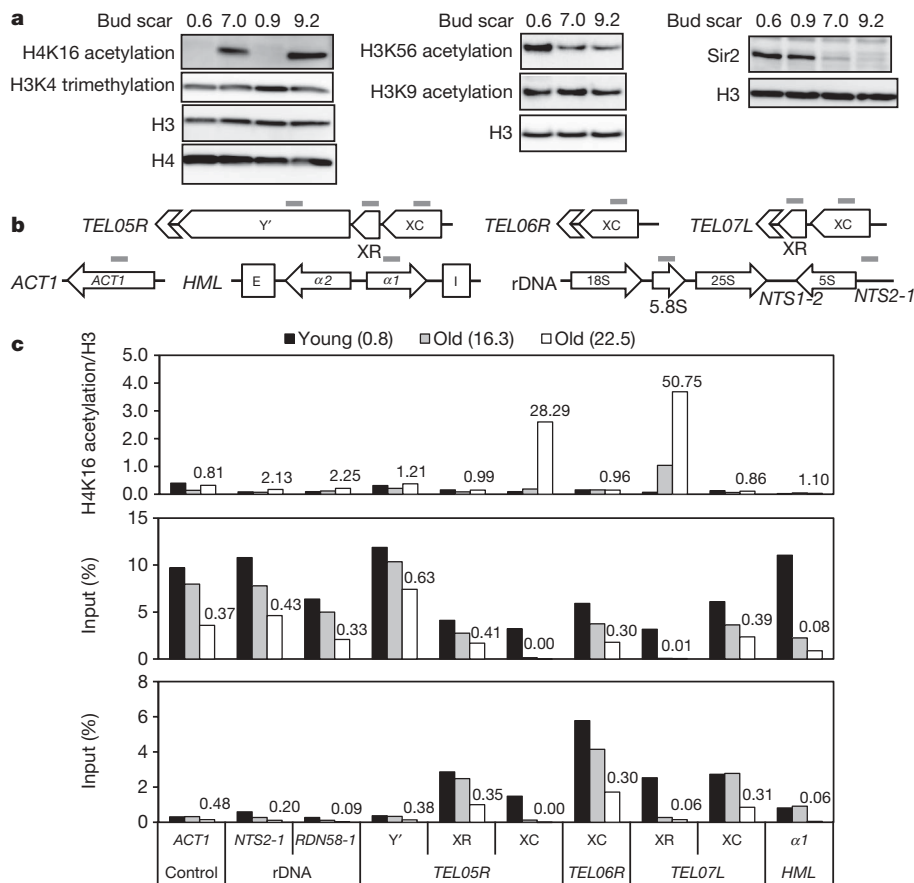


Figure 1 | Chromatin and Sir2 levels change in old cells. **a**, Western analysis for extracts from young and old yeast (strain W1588-4C) with specific antibodies. Consistent results were also observed with strain BY4741 (data not shown). **b**, Positions of quantitative PCR (qPCR) primers used in chromatin immunoprecipitation (ChIP). E, I, silencers; $\alpha 1$, *HML $\alpha 1$* ; $\alpha 2$, *HML $\alpha 2$* . **c**, ChIP analysis for young and old cells (average bud scar counts in

parentheses). Chromatin immunoprecipitated with antibodies against H4K16 acetylation (upper panel), H3 (middle panel) or Sir2 (lower panel) was analysed by qPCR with primers shown in **b**. Fold changes for the oldest sample (white bars) over the young cells (black bars) are indicated above the bars.

H4K16 and H3K56^{9,25}, and is linked to ageing²⁶; the abundance of Sir2 protein in age-sorted cells decreased as H4K16 acetylation increased (Fig. 1a, right panel). *SIR2* RNA levels remained unchanged in old cells (Supplementary Fig. 1c), as found previously²⁷, indicating that age-dependent changes in Sir2 occur post-transcriptionally.

Chromatin changes localize to silenced regions

To investigate the genomic location of increasing H4K16 acetylation and decreasing Sir2 abundance in old cells, we performed ChIP focusing on Sir2-regulated regions, including rDNA, telomeres and silenced mating-type loci (Fig. 1b). These regions, similar to heterochromatin in higher eukaryotes, have low levels of histone acetylation and minimal gene activity²⁸. Relative to young cells, we observed significantly higher levels of H4K16 acetylation in old cells at X core and X repeat elements within telomeres *TEL05R* and *TEL07L*, but not at other Sir2-regulated sites or control sites (Fig. 1c, upper panel). These hyperacetylated K16 sites co-localized with profoundly reduced histone levels in old cells, whereas histone levels were decreased to the same, moderate, extent at all other sites tested (Fig. 1c, middle panel). Consistent with the western analysis, Sir2 levels decreased at all Sir2-regulated sites in old cells (Fig. 1c, lower panel). We examined six additional telomeres and found that across nine total telomeres the greatest age-associated increase in H4K16 acetylation was within the X elements, correlating with the strongest decrease in histone levels (Supplementary Table 1). Because Sir2 was equally lost from all its binding sites, these results suggest that the chromatin landscape was mostly affected at X elements of telomeres as a result of loss of Sir2 in old cells.

To determine whether the observed age-associated chromatin changes lead to altered silencing state, we assayed the expression of a *URA3* reporter inserted at various positions near telomere *TEL11L*²⁹ (Fig. 2a). Silencing of *URA3* was measured in young and old cells, by quantifying survival on media containing 5-fluoro-orotic acid (5-FOA), which is converted into the toxic 5-fluorouracil by Ura3. Silencing at the X core element, but not elsewhere near the telomere, was significantly reduced in old cells (Fig. 2b), and sensitivity to 5-FOA was not seen in similarly aged *ura3 Δ* cells (Supplementary Fig. 4).

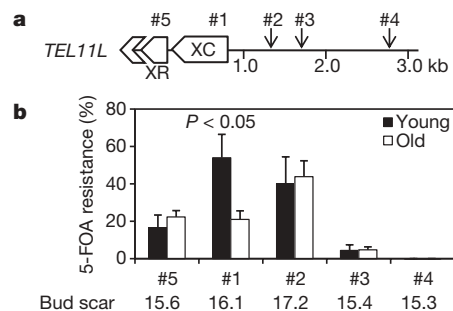


Figure 2 | The X core elements of telomeres show silencing defects in old cells. **a**, The positions (#1 to #5) of *URA3* gene insertions near telomere *TEL11L*²⁹. kb, kilobase. **b**, Silencing assay for young and old cells of strains bearing *URA3* gene insertions as shown in **a**. The extent of silencing is expressed as the fraction of cells resistant to 5-FOA ($n = 4$; error bars, s.d.). Average bud scar counts are listed for old-cell samples.

These data demonstrate that the increase in H4K16 acetylation, decrease in histone abundance and reduced Sir2 abundance in ageing cells are associated with transcriptional derepression at specific loci near telomeres.

Because ageing leads to decreased Sir2 abundance, we predicted that, in young cells, loss or inhibition of Sir2 might increase H4K16 acetylation and reduce histones at telomeric X elements. As seen previously³⁰, H4K16 acetylation increased at Sir2-regulated sites in *sir2Δ* cells (Supplementary Fig. 5a); as seen in old wild-type cells, young *sir2Δ* cells showed decreased histones at X elements in telomeres, but not at rDNA, HML or other control sites (Supplementary Fig. 5b). Similar results were observed for cells treated with nicotinamide (Supplementary Fig. 5c–e), a non-competitive inhibitor of Sir2³¹.

Sas2 antagonizes Sir2 effects on chromatin and lifespan

Sas2 is the major H4K16 acetyltransferase to establish boundaries between telomeres and euchromatin^{10,11}; hence, we conjectured that deletion of *SAS2* might extend lifespan. Indeed, the mean lifespan of *sas2Δ* was greater than wild type (Fig. 3a). Overexpression of *SIR2* by integrating a second copy also extends lifespan¹³. If this extension is due to enhanced deacetylation of H4K16, then *SIR2* overexpression should not further extend the lifespan of *sas2Δ* cells. Consistent with this prediction, we found that *sas2Δ SIR2*-overexpression cells were not longer lived than either *SIR2*-overexpression or *sas2Δ* cells (Fig. 3b).

Upon loss of Sir2, Sas2 spreads outward to telomeres to acetylate H4K16, resulting in disruption of telomere silencing¹⁰. Because deletion of *SAS2* increases lifespan and decreases H4K16 acetylation at

subtelomeric regions³², we speculated that Sas2 modulates chromatin changes in old cells. Indeed, relative to wild type, *sas2Δ* cells did not show the profound age-associated increase in H4K16 acetylation, decrease in histones or loss of Sir2 (Fig. 3c–e). Western-blot analysis confirmed that deletion of *SAS2* reduced overall abundance of H4K16 acetylation and stabilized histones and Sir2 in aged cells (Fig. 3f). Taken together, these data suggest that Sir2 and Sas2 antagonistically modulate lifespan in part through regulating H4K16 acetylation and histone levels at telomeres, and further suggest that Sas2 promotes degradation of Sir2 in old cells through unknown mechanisms.

H4K16 and H3K56 mutations result in shortening of lifespan

Our observations implicate H4K16 acetylation as a key substrate of Sir2 in opposing replicative ageing. To investigate the role of H4K16 in ageing more directly, we created strains bearing substitutions of H4K16 to arginine (K16R) or glutamine (K16Q), respectively mimicking the unacetylated or acetylated state. We found that K16Q significantly reduced lifespan, whereas K16R showed only a marginal effect (Fig. 4a). Neither substitution affected growth (Supplementary Fig. 10), and similar results were observed in a long-lived strain background carrying integrated histone mutations (Fig. 5a). It is notable that K16R did not extend lifespan. However, an arginine substitution may not perfectly mimic unacetylated lysine, and global K16R substitution may not have the opposite effect of global K16Q mutation, because H4K16 acetylation could have additional roles in regulating euchromatic genes important for normal lifespan³³. Sir2

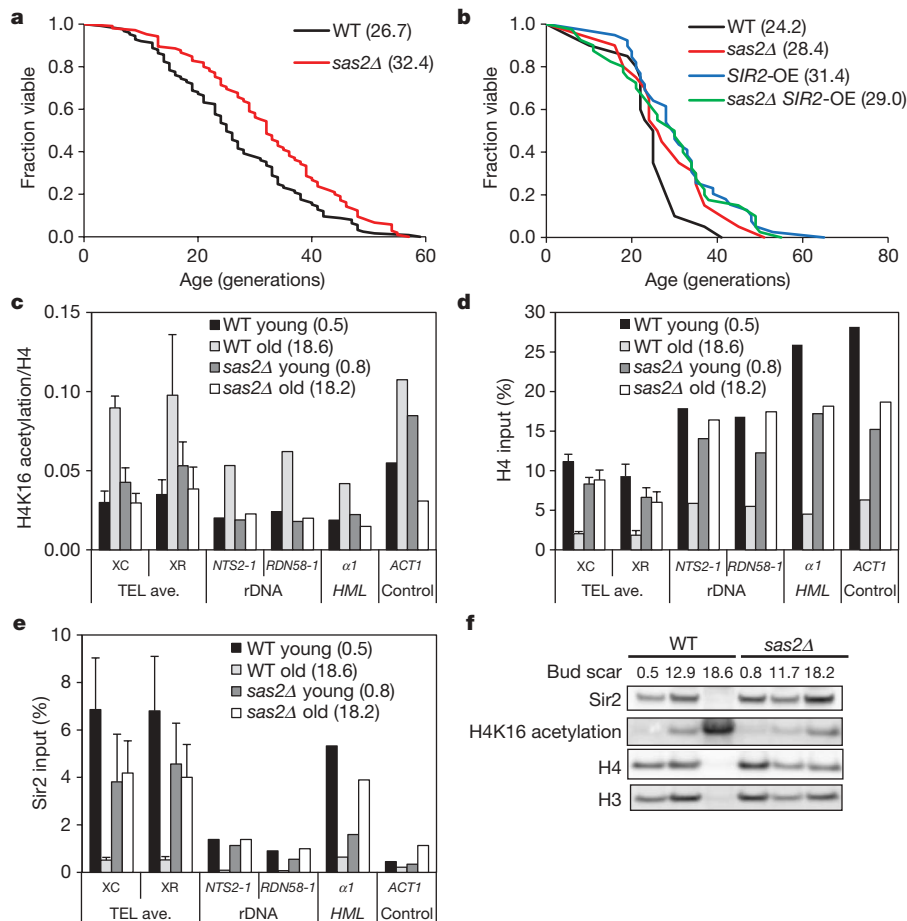


Figure 3 | Sas2 antagonizes Sir2 in regulating lifespan and H4K16 acetylation. **a, b**, Replicative lifespan analysis for strains WT (wild type) and *sas2Δ* (**a**) and strains WT, *sas2Δ*, *SIR2*-overexpression (*SIR2*-OE) and *sas2Δ SIR2*-overexpression (**b**) with mean lifespan (MLS) in parentheses. Lifespan *P* values are listed in Supplementary Table 5. **c–e**, Old cells were isolated for WT and *sas2Δ* strains (average bud scar counts in parentheses). Chromatin

immunoprecipitated with antibodies specific to H4K16 acetylation (**c**), H4 (**d**) or Sir2 (**e**) was analysed by qPCR with primers used in Supplementary Table 1. Data for X core (XC) positions and X repeat (XR) positions were averaged; error bars, s.d. **f**, Cellular extracts as in **c–e** were analysed by western blotting, probed with specific antibodies. TEL, telomere.

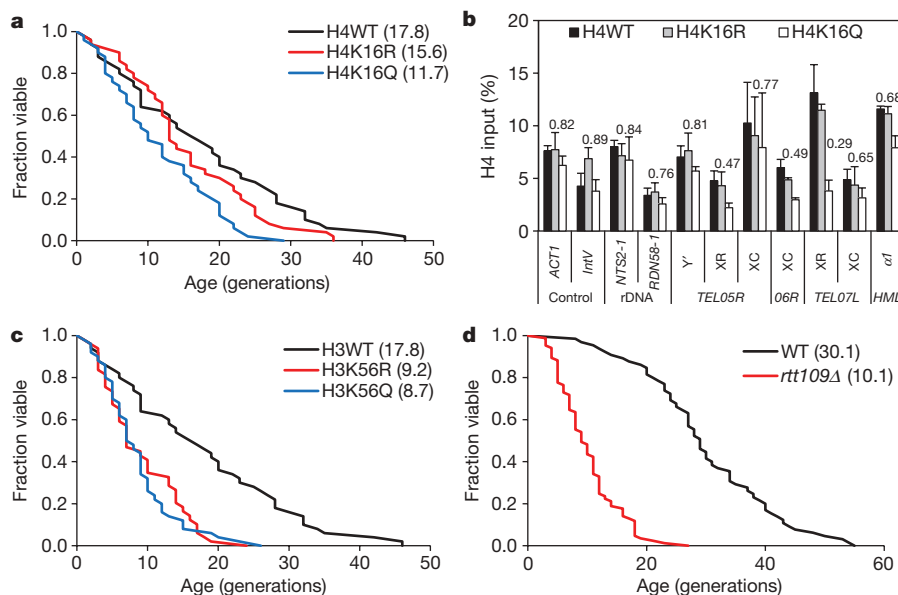


Figure 4 | Histone H4K16 and H3K56 mutations affect replicative lifespan through distinct mechanisms. **a, b**, Replicative lifespan analysis (MLS in parentheses) (**a**) and ChIP analysis with histone H4 antibody (**b**) for cells bearing WT histone H4, H4K16R or H4K16Q plasmids. Fold changes for H4K16Q compared with H4WT in ChIP analysis are indicated above the

abundance was unchanged in the mutant strains (Supplementary Fig. 6), indicating that the reduced lifespan is not due to loss of Sir2. We found lowered histone (Fig. 4b) and Sir2 (Supplementary Fig. 7) content at telomeres only in the H4K16Q strain, consistent with H4K16 hyperacetylation directly causing histone loss at telomeres.

bars ($n = 3$; error bars, s.d.). *IntV*, intergenic region on chromosome V (Supplementary Information). **c, d**, Replicative lifespan analysis for strains carrying WT histone H3, H3K56R or H3K56Q plasmids (**c**) and strains WT or *rtt109Δ* (**d**), with MLS in parentheses. Lifespan *P* values are listed in Supplementary Table 5.

H3K56 acetylation, regulated by two other sirtuins, Hst3 and Hst4³⁴, and important for genome integrity and replication^{34–36}, was found in decreased abundance in old cells (Fig. 1a). In contrast to H4K16, both H3K56R and H3K56Q exhibited similarly reduced lifespans (Fig. 4c). H3K9 substitutions had no effect on lifespan (Supplementary Fig. 8).

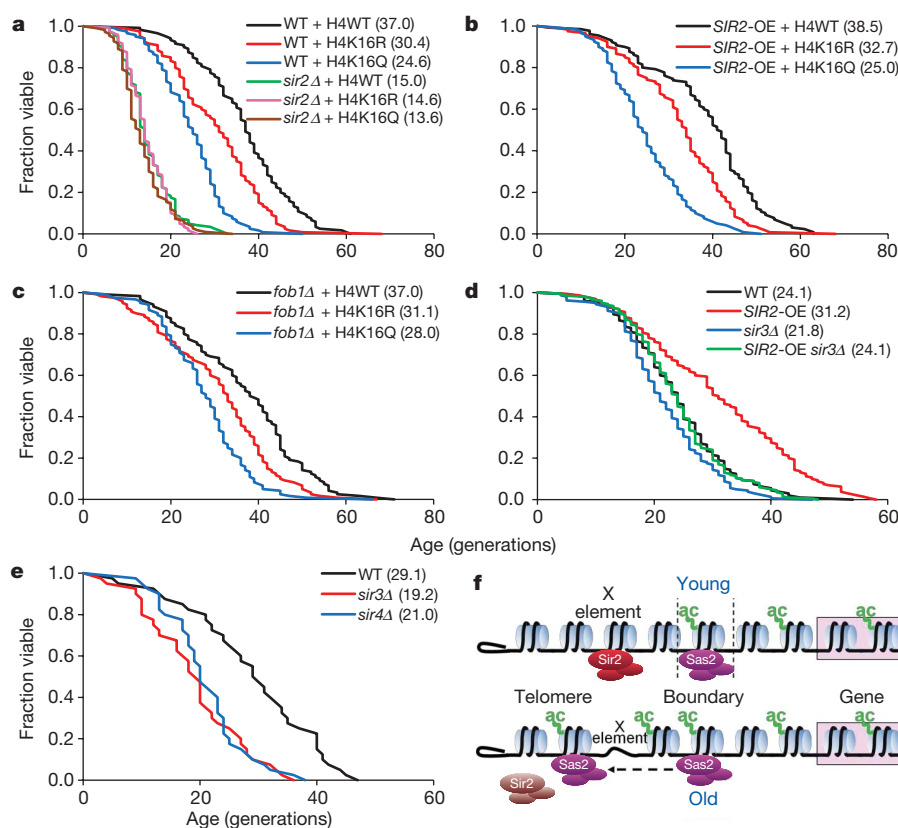


Figure 5 | H4K16 is involved in a Sir2-regulated ageing pathway associated with telomere chromatin. **a–c**, Replicative lifespan analysis performed in parallel for strains bearing integrated WT histone H4, H4K16R or H4K16Q in WT and *sir2Δ* backgrounds (**a**), in *SIR2*-overexpression background (**b**) and in *fob1Δ* background (**c**), with MLS in parentheses. **d, e**, Replicative

lifespan analysis for strains WT, *SIR2*-overexpression, *sir3Δ* and *SIR2*-overexpression *sir3Δ* (**d**), and for strains WT, *sir3Δ* and *sir4Δ* (**e**), with MLS in parentheses. Lifespan *P* values are listed in Supplementary Table 5. **f**, Model for telomere chromatin changes in yeast old mother cells. ac, acetylation.

Deletion of *RTT109*, which expresses the key acetyltransferase for H3K56³⁷, also reduced lifespan (Fig. 4d). Finally, as shown previously³⁸, *hst3Δ*, *hst4Δ* or *hst3Δ hst4Δ* cells had a significantly decreased lifespan (Supplementary Fig. 9). H3K56R substitution or *rtt109Δ* causes sensitivity to DNA damage^{35,36}, suggesting that acetylation of K56 is required for genome integrity. Because K56Q, *hst3Δ* and *hst3Δ hst4Δ* showed similar shortening of lifespan in comparison with K56R and *rtt109Δ*, we reasoned that cycling of acetylation and deacetylation of K56 might be important for genome integrity. K56Q, but not H4K16 or H3K9, was equally as sensitive to methyl methanesulphonate and hydrogen peroxide as K56R (Supplementary Fig. 10). We conclude that acetylation of H4K16 and H3K56 both influence longevity, but through distinct mechanisms.

Shortening of lifespan by H4K16 mutants is epistatic with Sir2

Deletion of the WRN homologue *SGS1* shortens lifespan^{12,13}. However, in contrast to *sir2Δ*, *sgs1Δ* did not affect H4K16 acetylation or histone levels (Supplementary Fig. 11). To investigate the genetic pathways of H4K16 acetylation, we compared the lifespans of H4K16 substitutions in *sir2Δ* and *sgs1Δ* backgrounds. H4K16 mutations did not further shorten the lifespans of *sir2Δ* cells (Fig. 5a), indicating that K16 functions in the same pathway as Sir2. In contrast, the K16 substitutions significantly shortened the lifespans of *sgs1Δ* cells (Supplementary Fig. 12), suggesting that K16 and Sir2 act in a different pathway from *Sgs1*. We also examined whether H4K16 is critical for lifespan extension by *SIR2* overexpression, and found that K16Q in the *SIR2*-overexpression background restricted lifespan in a manner similar to K16Q in wild type (compare Fig. 5b with Fig. 5a). Thus, hypoacetylated H4K16 is required for the lifespan extension by *SIR2* overexpression. Overall, these data support the idea that Sir2 deacetylates K16 to antagonize ageing.

H4K16 functions in a pathway distinct from ERC formation

Our data show age-related Sir2-linked chromatin changes at telomeres, but previous findings have indicated that Sir2 promotes replicative longevity by functioning at the rDNA to inhibit the formation of extrachromosomal rDNA circles, which accumulate with age in mother cells and cause senescence¹³. Fob1 acts antagonistically to Sir2 by promoting rDNA recombination, and deletion of *FOB1* reduces the number of ERCs and extends lifespan¹⁵. We reasoned that, if K16Q shortens lifespan through effects exclusively at rDNA, then K16Q will not shorten lifespan in a *fob1Δ* background. However, K16Q did reduce the lifespan of *fob1Δ* cells, and this reduction was greater than the effect of K16R in combination with *fob1Δ* (Fig. 5c). This result is similar to the shorter lifespan of *sir2Δ fob1Δ* in comparison with *fob1Δ* alone¹³, suggesting that K16Q influences lifespan through a mechanism different from Fob1 and ERCs.

Sir2 forms a complex with Sir3 and Sir4 at telomeres and silenced mating-type loci, and a distinct complex, RENT, that functions at rDNA³⁹. We found that the lifespan extension by *SIR2* overexpression is suppressed by deletion of *SIR3* (Fig. 5d) and that deletion of either *SIR3* or *SIR4* similarly shortens lifespan (Fig. 5e). Taken together, these data support the idea that the Sir2–Sir3–Sir4 complex functions to modulate ageing by maintaining telomeric chromatin, and that H4K16 acetylation is a key substrate of Sir2 during replicative ageing. Furthermore, our data suggest that the age-associated chromatin alterations at telomeres and other silenced loci results from decreased abundance of Sir2 protein during ageing, which can be counteracted by overexpression of *SIR2*.

Discussion

Saccharomyces cerevisiae provides a previously unexploited means of analysing chromatin during ageing. Our study identifies a novel role for H4K16 acetylation and telomere chromatin state in determining longevity, mediated by Sas2 acetylation and Sir2 deacetylation. We show an age-associated decrease in Sir2 protein abundance accompanied by an increase in H4K16 acetylation and loss of histones at

specific subtelomeric regions in replicatively old cells (Fig. 5f), which results in compromised transcriptional silencing at these loci. Deletion of *SAS2* stabilizes Sir2 levels in old cells and extends lifespan. Finally, we demonstrate that mutations to H4K16 negatively affect lifespan, downstream of Sir2, in a pathway at least partly distinct from the accumulation of ERCs in old cells.

Although extensive evidence supports the idea that replicative lifespan is regulated by Sir2 function at rDNA, a Sir2-related ageing mechanism distinct from ERC accumulation has been suggested^{26,40}. Our data indicate that Sir2 also promotes replicative longevity through at least one additional pathway, apparently localized at telomeres, as depicted in Fig. 5f. In this pathway, Sir2 in young cells maintains low H4K16 acetylation at telomeres and subtelomeric regions and works with Sas2 to establish a silencing boundary^{10,11}. In old cells, owing to loss of Sir2 and action of Sas2, H4K16 acetylation increases leading to loss of histones at X elements of telomeres. A role for Sir2 in counteracting H4K16 acetylation specifically in a telomere-mediated ageing pathway is potentially evolutionarily conserved, as telomeric changes have a major role in ageing and cancer and correlate with chromatin alterations⁴¹. However, our data do not rule out non-telomeric location(s) that could be regulated through H4K16 acetylation by Sas2 and deacetylation by the SIR complex.

There are seven mammalian sirtuins, and three (SIRT1, SIRT2 and SIRT3) deacetylate H4K16²³. It is possible that reduction of activity or abundance of these sirtuins promotes cellular ageing through increased H4K16 acetylation. SIRT6 regulates lifespan through deacetylation of H3K9ac^{22,42}, and SIRT6 knockout mice show genomic instability, including telomeric fusions, and an ageing-like phenotype⁴³. Hence, regulation of histone acetylation appears to be a critical function in ageing pathways throughout eukaryotes.

METHODS SUMMARY

Yeast old mother cells were isolated by 2–4 rounds of sorting; cells were labelled with biotin, cultured overnight, and this was followed by affinity purification²⁴. Mean ages of isolated cells were estimated by counting Calcofluor-stained bud scars. Whole-cell extracts of young and old cells were analysed by western blotting with specific antibodies. Specificity of key antibodies (H4K16 acetylation, H3K56 acetylation, H3K9 acetylation, H3K4 trimethylation and Sir2) used in this study was verified by western blotting with extracts from cells bearing corresponding histone point mutations or gene deletions. Total DNA and messenger RNA extractions followed by real-time PCR were used to quantify rDNA copy numbers and expression levels of *SIR2*. Formaldehyde-crosslinked young and old cells were used for ChIP with specific antibodies. Immunoprecipitated DNA was quantified by real-time PCR with primer sets targeting specific regions. Telomere silencing assays were performed with strains and methods described elsewhere²⁹. Replicative lifespans of yeast strains were determined as described¹³. Statistical assessment of lifespan differences was performed using the Wilcoxon rank-sum test. Yeast strains bearing histone point mutations were generated by deletion of both copies of histone H3 and H4 gene cassettes supplemented by a wild-type or mutant copy of H3 and H4 either on a *CEN6*-controlled yeast plasmid or by integrating a copy to its original gene location. Strains, antibodies and real-time PCR oligonucleotides used in this study are listed in Supplementary Tables 2, 3, and 4, respectively.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 February; accepted 21 April 2009.

- Collado, M., Blasco, M. A. & Serrano, M. Cellular senescence in cancer and aging. *Cell* **130**, 223–233 (2007).
- Herbig, U. et al. Cellular senescence in aging primates. *Science* **311**, 1257 (2006).
- Mair, W. & Dillin, A. Aging and survival: the genetics of life span extension by dietary restriction. *Annu. Rev. Biochem.* **77**, 727–754 (2008).
- Guarente, L. & Kenyon, C. Genetic pathways that regulate ageing in model organisms. *Nature* **408**, 255–262 (2000).
- Fraga, M. F. & Esteller, M. Epigenetics and aging: the targets and the marks. *Trends Genet.* **23**, 413–418 (2007).
- Oberdoerffer, P. & Sinclair, D. A. The role of nuclear architecture in genomic instability and ageing. *Nature Rev. Mol. Cell Biol.* **8**, 692–702 (2007).

7. Krishnamoorthy, T. *et al.* Phosphorylation of histone H4 Ser1 regulates sporulation in yeast and is conserved in fly and mouse spermatogenesis. *Genes Dev.* **20**, 2580–2592 (2006).
8. Longo, V. D. & Kennedy, B. K. Sirtuins in aging and age-related disease. *Cell* **126**, 257–268 (2006).
9. Moazed, D. Enzymatic activities of Sir2 and chromatin silencing. *Curr. Opin. Cell Biol.* **13**, 232–238 (2001).
10. Kimura, A., Umehara, T. & Horikoshi, M. Chromosomal gradient of histone acetylation established by Sas2p and Sir2p functions as a shield against gene silencing. *Nature Genet.* **32**, 370–377 (2002).
11. Suka, N., Luo, K. & Grunstein, M. Sir2p and Sas2p opposingly regulate acetylation of yeast histone H4 lysine16 and spreading of heterochromatin. *Nature Genet.* **32**, 378–383 (2002).
12. Sinclair, D. A. & Guarente, L. Extrachromosomal rDNA circles—a cause of aging in yeast. *Cell* **91**, 1033–1042 (1997).
13. Kaerberlein, M., McVey, M. & Guarente, L. The SIR2/3/4 complex and SIR2 alone promote longevity in *Saccharomyces cerevisiae* by two different mechanisms. *Genes Dev.* **13**, 2570–2580 (1999).
14. Johzuka, K. & Horiuchi, T. Replication fork block protein, Fob1, acts as an rDNA region specific recombinator in *S. cerevisiae*. *Genes Cells* **7**, 99–113 (2002).
15. Defossez, P. A. *et al.* Elimination of replication block protein Fob1 extends the life span of yeast mother cells. *Mol. Cell* **3**, 447–455 (1999).
16. Johnson, F. B., Sinclair, D. A. & Guarente, L. Molecular biology of aging. *Cell* **96**, 291–302 (1999).
17. Laun, P. *et al.* Yeast mother cell-specific ageing, genetic (in)stability, and the somatic mutation theory of ageing. *Nucleic Acids Res.* **35**, 7514–7526 (2007).
18. Lin, Y. Y. *et al.* Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis. *Cell* **136**, 1073–1084 (2009).
19. Kouzarides, T. Acetylation: a regulatory modification to rival phosphorylation? *EMBO J.* **19**, 1176–1179 (2000).
20. Glazak, M. A. & Seto, E. Histone deacetylases and cancer. *Oncogene* **26**, 5420–5432 (2007).
21. Haigis, M. C. & Guarente, L. P. Mammalian sirtuins—emerging roles in physiology, aging, and calorie restriction. *Genes Dev.* **20**, 2913–2921 (2006).
22. Michishita, E. *et al.* SIRT6 is a histone H3 lysine 9 deacetylase that modulates telomeric chromatin. *Nature* **452**, 492–496 (2008).
23. Saunders, L. R. & Verdin, E. Sirtuins: critical regulators at the crossroads between cancer and aging. *Oncogene* **26**, 5489–5504 (2007).
24. Smeal, T. *et al.* Loss of transcriptional silencing causes sterility in old mother cells of *S. cerevisiae*. *Cell* **84**, 633–642 (1996).
25. Xu, F. *et al.* Sir2 deacetylates histone H3 lysine 56 to regulate telomeric heterochromatin structure in yeast. *Mol. Cell* **27**, 890–900 (2007).
26. Kaerberlein, M. & Powers, R. W. 3rd Sir2 and calorie restriction in yeast: a skeptical perspective. *Ageing Res. Rev.* **6**, 128–140 (2007).
27. Lesur, I. & Campbell, J. L. The transcriptome of prematurely aging yeast cells is similar to that of telomerase-deficient cells. *Mol. Biol. Cell* **15**, 1297–1312 (2004).
28. Rusche, L. N., Kirchmaier, A. L. & Rine, J. The establishment, inheritance, and function of silenced chromatin in *Saccharomyces cerevisiae*. *Annu. Rev. Biochem.* **72**, 481–516 (2003).
29. Pryde, F. E. & Louis, E. J. Limitations of silencing at native yeast telomeres. *EMBO J.* **18**, 2538–2550 (1999).
30. Robyr, D. *et al.* Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell* **109**, 437–446 (2002).
31. Grubisha, O., Smith, B. C. & Denu, J. M. Small molecule regulation of Sir2 protein deacetylases. *FEBS J.* **272**, 4607–4616 (2005).
32. Shia, W. J., Li, B. & Workman, J. L. SAS-mediated acetylation of histone H4 Lys 16 is required for H2A.Z incorporation at subtelomeric regions in *Saccharomyces cerevisiae*. *Genes Dev.* **20**, 2507–2512 (2006).
33. Dion, M. F., Altschuler, S. J., Wu, L. F. & Rando, O. J. Genomic characterization reveals a simple histone H4 acetylation code. *Proc. Natl Acad. Sci. USA* **102**, 5501–5506 (2005).
34. Celic, I. *et al.* The sirtuins hst3 and Hst4p preserve genome integrity by controlling histone h3 lysine 56 deacetylation. *Curr. Biol.* **16**, 1280–1289 (2006).
35. Driscoll, R., Hudson, A. & Jackson, S. P. Yeast Rtt109 promotes genome stability by acetylating histone H3 on lysine 56. *Science* **315**, 649–652 (2007).
36. Han, J. *et al.* Rtt109 acetylates histone H3 lysine 56 and functions in DNA replication. *Science* **315**, 653–655 (2007).
37. Schneider, J. *et al.* Rtt109 is required for proper H3K56 acetylation: a chromatin mark associated with the elongating RNA polymerase II. *J. Biol. Chem.* **281**, 37270–37274 (2006).
38. Tsuchiya, M. *et al.* Sirtuin-independent effects of nicotinamide on lifespan extension from calorie restriction in yeast. *Ageing Cell* **5**, 505–514 (2006).
39. Huang, J. & Moazed, D. Association of the RENT complex with nontranscribed and coding regions of rDNA and a regional requirement for the replication fork block protein Fob1 in rDNA silencing. *Genes Dev.* **17**, 2162–2176 (2003).
40. Aguilaniu, H., Gustafsson, L., Rigoulet, M. & Nystrom, T. Asymmetric inheritance of oxidatively damaged proteins during cytokinesis. *Science* **299**, 1751–1753 (2003).
41. Blasco, M. A. The epigenetic regulation of mammalian telomeres. *Nature Rev. Genet.* **8**, 299–309 (2007).
42. Kawahara, T. L. *et al.* SIRT6 links histone H3 lysine 9 deacetylation to NF-kappaB-dependent gene expression and organismal life span. *Cell* **136**, 62–74 (2009).
43. Mostoslavsky, R. *et al.* Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. *Cell* **124**, 315–329 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Louis for providing strains for the telomere silencing assay, and members of the Kennedy and Kaerberlein labs who participated in lifespan-determination studies. This work was funded by US National Institutes of Health grants (S.L.B. and B.K.K.), and an American Federation for Aging Research Julie Martin Mid-Career Award in Aging Research (B.K.K.). M.K. is an Ellison Medical Foundation New Scholar in Aging.

Author Contributions Project planning was performed by W.D., F.B.J., A.S., B.K.K. and S.L.B.; experimental work by W.D., K.K.S., R.P. and J.A.D.; data analysis by W.D., K.K.S., F.B.J., M.K., B.K.K. and S.L.B.; and manuscript composition by W.D., F.B.J., M.K., B.K.K. and S.L.B.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.L.B. (sberger@mail.med.upenn.edu).

METHODS

Plasmids and strains. Plasmids pWD23, pWD25, pWD35, pWD36, pWD43 and pWD45 bearing point mutations of H4K16R, H4K16Q, H3K9R, H3K9Q, H3K56R and H3K56Q, respectively, were generated by QuikChange site-directed mutagenesis (Stratagene) from plasmid pRM204 containing a wild-type copy of the histone H3 and H4 genes (*HHT2–HHF2*) and verified by sequencing. Yeast strains used in this study are listed in Supplementary Table 2. Strains carrying a mutant histone plasmid as the sole source of histones H3 and H4 were made by shuffling the mutant plasmid into strain FY1716, followed by selection on synthetic complete media containing 1 mg ml^{-1} 5-FOA to remove the existing wild-type histone plasmid pDM1 from FY1716. Strains bearing an integrated histone mutation were constructed by transforming PCR products containing the mutation into strain FY1716 and selecting on media containing 5-FOA. All histone mutant strains were verified by sequencing.

Isolation of yeast old mother cells. Old mother cells were isolated from exponentially growing cultures in YPD (1% yeast extract, 2% Bacto peptone, 2% dextrose) by surface labelling with NHS-LC-Biotin (Thermo Fisher) and affinity purification as previously described⁴⁴, except that up to 4.8×10^9 cells were used for labelling for up to four rounds of sorting and Dynabeads Biotin Binder (Invitrogen) was used. Exponential growth before each round of sorting was limited to 6–8 doublings. About 4×10^8 old mother cells were saved for analysis after each round of sorting. Mean ages of isolated cells ($n = 50$ each) were estimated by counting the number of bud scars, stained with Calcofluor (Sigma) and visualized by fluorescence microscopy.

Preparation of whole-cell extracts and western blotting. Whole-cell extracts were prepared as described⁷. Proteins were resolved by SDS–polyacrylamide gel electrophoresis and transferred to polyvinylidene difluoride. Blots were probed with antibodies specific for histones H3 and H4, various histone modifications and Sir2, as listed in Supplementary Table 3.

Quantification of ribosomal DNA copy number and messenger RNA expression levels. Total DNA was purified typically from 100- μg whole-cell extracts of

young or old cells using a QIAGEN PCR purification kit following sequential 10- μg RNase A treatment and 20- μg protease K treatment at 37°C for 1 h each. Levels of rDNA were quantified by real-time PCR with primers specific to the 5.8S rDNA gene *RDN58-1* and non-transcribed region *NTS2-1*, normalized to *ACT1* levels. Total mRNA and complementary DNA were prepared using a QIAGEN RNeasy kit and an Applied Biosystems reverse transcription kit, and quantified by real-time PCR. Real-time PCR primers are listed in Supplementary Table 4.

Chromatin immunoprecipitation. Young or old cell samples were crosslinked with 1% formaldehyde at room temperature for 10 min immediately after cell sorting or harvesting. ChIP was performed as described⁴⁵ with specific antibodies and quantified by real-time PCR. The specificity of all antibodies used in ChIP (listed in Supplementary Table 3) was verified using lysates prepared from strains carrying corresponding histone substitution mutations or gene deletions. Real-time PCR primers are listed in Supplementary Table 4.

Telomere silencing. Young and old cells were isolated as described above except that the scale was reduced 20-fold. About 1×10^7 young and old cells were resuspended in 200 μl , followed by 10-fold serial dilutions. Then 10 μl of each dilution was spotted on fresh synthetic complete media and synthetic complete media containing 1 mg ml^{-1} 5-FOA. Fractions of cells giving rise to 5-FOA resistant colonies are shown.

Yeast lifespan determination. Replicative lifespans of yeast strains were determined by micromanipulation as described¹³. Statistical assessment of lifespan differences was determined using the Wilcoxon rank-sum test.

Nicotinamide treatment. Yeast cells were grown at 30°C overnight in YPD, then transferred to fresh YPD or YPD containing 5 mM nicotinamide and allowed to continue growing for 4–5 h (two doublings of exponential growth).

44. Park, P. U., McVey, M. & Guarente, L. Separation of mother and daughter cells. *Methods Enzymol.* **351**, 468–477 (2002).

45. Wyce, A. *et al.* H2B ubiquitylation acts as a barrier to Ctk1 nucleosomal recruitment prior to removal by Ubp8 within a SAGA-related complex. *Mol. Cell* **27**, 275–288 (2007).

ARTICLES

Flipping of alkylated DNA damage bridges base and nucleotide excision repair

Julie L. Tubbs¹, Vitaly Latypov², Sreenivas Kanugula³, Amna Butt², Manana Melikishvili⁴, Rolf Kraehenbuehl⁵†, Oliver Fleck⁵†, Andrew Marriott², Amanda J. Watson², Barbara Verbeek²†, Gail McGown², Mary Thorncroft², Mauro F. Santibanez-Koref⁶, Christopher Millington⁷, Andrew S. Arvai¹, Matthew D. Kroeger¹, Lisa A. Peterson⁸, David M. Williams⁷, Michael G. Fried⁴, Geoffrey P. Margison², Anthony E. Pegg³ & John A. Tainer^{1,9}

Alkyltransferase-like proteins (ATLs) share functional motifs with the cancer chemotherapy target O⁶-alkylguanine-DNA alkyltransferase (AGT) and paradoxically protect cells from the biological effects of DNA alkylation damage, despite lacking the reactive cysteine and alkyltransferase activity of AGT. Here we determine *Schizosaccharomyces pombe* ATL structures without and with damaged DNA containing the endogenous lesion O⁶-methylguanine or cigarette-smoke-derived O⁶-4-(3-pyridyl)-4-oxobutylguanine. These results reveal non-enzymatic DNA nucleotide flipping plus increased DNA distortion and binding pocket size compared to AGT. Our analysis of lesion-binding site conservation identifies new ATLs in sea anemone and ancestral archaea, indicating that ATL interactions are ancestral to present-day repair pathways in all domains of life. Genetic connections to mammalian XPG (also known as ERCC5) and ERCC1 in *S. pombe* homologues Rad13 and Swi10 and biochemical interactions with *Escherichia coli* UvrA and UvrC combined with structural results reveal that ATLs sculpt alkylated DNA to create a genetic and structural intersection of base damage processing with nucleotide excision repair.

DNA O⁶-alkylguanine lesions are mutagenic and cytotoxic: they mispair during replication with thymine, resulting in G•C to A•T transition mutations^{1–4}. Human O⁶-alkylguanine DNA lesions are repaired by O⁶-alkylguanine-DNA alkyltransferase (AGT), also known as O⁶-methylguanine-DNA methyltransferase (MGMT) in humans, which transfers guanine O⁶-alkyl adducts to its reactive cysteine, reversing damage¹. This prevents mutations, but also provides resistance against alkylating chemotherapies^{2,5}. The active site PCHRV sequence motif Cys 145 (refs 6, 7) plus Arg 128 and Tyr 114 nucleotide rotating residues are conserved from bacterial to human AGTs^{1,2,8}. Human AGT structures alone^{9,10} and with small molecule⁹ or DNA^{11,12} substrates showed how AGT promotes resistance to anticancer therapies by directly reversing DNA guanine alkylation damage².

Recently, bacterial and yeast proteins with sequence similarity to the AGT DNA-binding domain were identified in which the cysteine alkyl acceptor was replaced by tryptophan, alanine or another residue¹³, and hence named alkyltransferase-like proteins (ATLs). ATLs from *S. pombe* (Atl1) and *E. coli* (eAtl) inhibit O⁶-methylguanine (O⁶-mG) repair by human AGT^{14,15}. *E. coli* Atl also binds abasic-site-containing double-stranded (ds)DNA¹⁶, and Atl1 binds single-stranded (ss)DNA containing O⁶-methyl-, O⁶-benzyl-, O⁶-(4-bromophenyl)- or O⁶-hydroxyethyl-guanine. However, ATLs do not cleave the alkyl group, base or oligonucleotide near the lesion^{14,15}, and the *E. coli* Atl Trp-to-Cys mutation does not restore alkyltransferase activity¹⁴. Because *S. pombe* and *Thermus thermophilus* lack AGT, and inactivation of their ATL genes (*atl1* and *TTHA1564*,

respectively^{15,17}) reduces their alkylation damage resistance, ATLs protect against biological effects of DNA alkylation damage by an undefined mechanism.

Tight binding affinities for^{16,17} and inability to repair^{14,15} O⁶-alkyl lesions implied ATLs are damage sensors or act in nucleotide excision repair (NER)^{13,15}, which excises bulky, DNA-distorting lesions. However, the lack of structures, persuasive evidence or a specific mechanism has obscured how ATL ameliorates DNA damage effects. To clarify this protection, we combined structural, biochemical and genetic experiments on Atl1 from the fission yeast *S. pombe*. Our results reveal that ATL binding generates a stable complex that sculpts alkylated DNA base damage for NER pathway entry.

Atl1 structure and lesion binding

To characterize Atl1–DNA damage interactions, we crystallized and solved structures to 2.0, 2.7 and 2.8 Å resolution, respectively, for Atl1 alone (Fig. 1a and Supplementary Table 1) and in complex with oligonucleotides containing either O⁶-mG (Fig. 1b, d and Supplementary Table 1) or O⁶-4-(3-pyridyl)-4-oxobutylguanine (O⁶-pobG; Fig. 1c and Supplementary Table 1), a bulky and toxicologically relevant adduct¹⁸. Atl1 shares the human AGT catalytic domain fold (superposition root mean squares difference of 1.6 Å; Fig. 1a), including residues required for AGT activity, DNA binding and nucleotide flipping. Yet, Atl1 specifically lacks the active site Cys and Asn hinge of AGT that couples helix–turn–helix DNA binding and active site motifs (Fig. 1a).

¹Skaggs Institute for Chemical Biology and Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037, USA. ²Cancer Research UK Carcinogenesis Group, Paterson Institute for Cancer Research, University of Manchester, Manchester, M20 4BX, UK. ³Department of Cellular and Molecular Physiology, Milton S. Ebersole Medical Center, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA. ⁴Center for Structural Biology, Department of Molecular and Cellular Biochemistry, University of Kentucky, Lexington, Kentucky 40536, USA. ⁵NWCRF Institute, Bangor University, Gwynedd LL57 2UW, UK. ⁶Institute of Human Genetics, Newcastle University, Newcastle-upon-Tyne, NE1 3BZ, UK. ⁷Centre for Chemical Biology, Department of Chemistry, University of Sheffield, Sheffield S3 7HF, UK. ⁸Division of Environmental Health Sciences and the Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455, USA. ⁹Life Sciences Division, Bioenergy and Structural Biology, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [†]Present addresses: Cancer Research UK DNA Damage Response Group, Paterson Institute for Cancer Research, University of Manchester, Manchester, UK (R.K.); Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark (O.F.); Department of Toxicology, University of Mainz, D-55131 Mainz, Germany (B.V.).

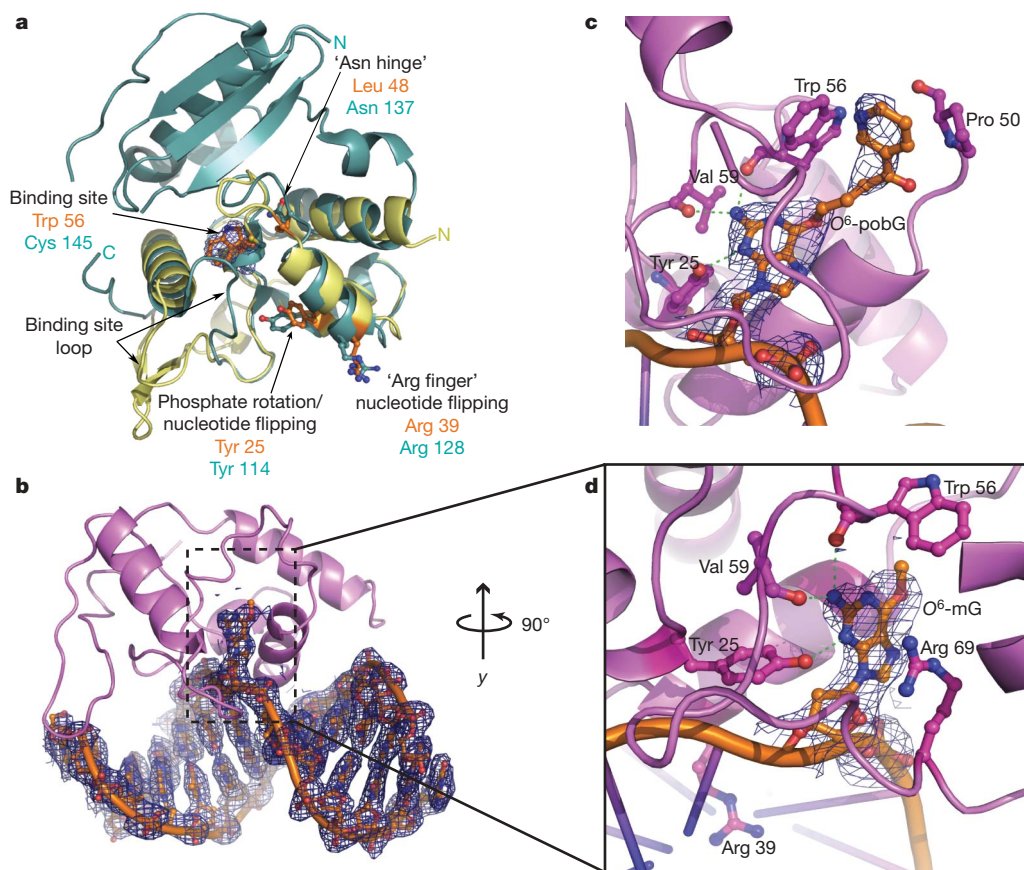


Figure 1 | Atl1 structure and lesion-binding site. **a**, Overlay of Atl1 (yellow) and AGT (cyan; PDB accession code 1EH6) models and comparison of key functional residues. $2F_o - F_c$ electron density (blue) for Atl1 with the binding site Trp 56 side chain omitted. **b**, Atl1 (magenta) bound to DNA containing O^6 -mG (orange). $2F_o - F_c$ simulated annealing composite omit map (blue) is shown for DNA. **c**, **d**, Atl1 lesion-binding site close-up with O^6 -pobG (**c**) or O^6 -mG (**d**). Amino acid side chains (ball-and-stick) and hydrogen bonds to the damaged guanine (green dashes) show the damage binding.

Atl1 flips both O^6 -mG (Fig. 1b, d) and O^6 -pobG (Fig. 1c) into a pocket containing PWHRV sequence motif Trp 56, consistent with fluorescence-measured flipping for a base opposite an AP site in *E. coli* Atl (ref. 16) and for O^6 -mG in TTHA1564 (ref. 17). Atl1 displays no AGT activity¹⁵, indicating that nucleotide flipping is a switch for pathway activation, not catalysis. To our knowledge, Atl1 (Fig. 1b–d), *E. coli* Atl (ref. 16) and TTHA1564 (ref. 17) are among the first reported non-enzymatic DNA-binding proteins that flip nucleotides into a specificity pocket. Arg 39 intercalates the DNA base stack (Fig. 1d) and hydrogen bonds with the orphaned cytosine, thereby stabilizing the extra-helical alkylguanine. Trp 56, rather than AGT Cys, is evident in electron density omit maps (Fig. 1a) and acts in hydrophobic packing with the alkyl group (Fig. 1c, d). Arg 69 guanidinium stacks against the alkylguanine base in a cation– π interaction (Fig. 1d).

Alkylguanine base side and main chain hydrogen bonds are conserved from Atl1 to AGT, but the Atl1 lesion-binding pocket is approximately three times larger (Supplementary Table 2). Loop residues 65–73 define one wall of the alkyl-binding pocket, adopting a conformation further from the protein core than in AGT, thereby enlarging the pocket (Figs 1a and 2a). Also, the lesion-binding pocket Lys 45–Pro 55 cap is ~ 5.3 Å further out than the comparable AGT Pro 140 ($C\alpha$ to $C\alpha$ distance), which interacts with larger alkyl groups^{9,11,12}. Moreover, ATL Ile 71 replaces AGT Tyr 158, which would clash with the Atl1 Trp 56 side chain in its DNA-bound, closed position.

The larger cavity of Atl1 explains its broad lesion range, which includes O^6 -benzyl-, O^6 -(4-bromophenyl)- or O^6 -hydroxyethyl-guanine. In the O^6 -pobG–DNA complex structure, the pob group is wedged between Pro 50 and Trp 56, making only these hydrophobic protein interactions (Fig. 1c). No major changes in lesion-binding site or DNA conformation occur between O^6 -mG- and O^6 -pobG-bound Atl1. Pob would push against the active (or binding)

site loop in the smaller AGT active site, explaining why AGT repairs pob lesions at a decreased rate compared to O^6 -mG¹⁹. Pob adopts a conformation incompatible with smaller *E. coli* AGT (Ada and Ogt) active site pockets, consistent with its poor repair by Ada and Ogt²⁰ and need for ATLs for bulky adducts in organisms such as *E. coli*.

DNA binding by Atl1

Like AGT, Atl1 uses a helix–turn–helix motif to bind the DNA minor groove (Figs 1b and 2a). All damaged-strand contacts are to alkyl-guanine and two 3'-adjacent nucleotide phosphate groups (Fig. 2d). DNA binding site loop (Ser 67 and Lys 70) and C-terminal loop (Thr 92 and Ser 93) residues form DNA contacts not found in AGT. DNA binding of Atl1 buries $\sim 1,050$ Å² versus 788 Å² of AGT buried surface area, consistent with tighter DNA binding^{16,17}.

Atl1 bends DNA by $\sim 45^\circ$ (Fig. 2a), whereas AGT only bends DNA by $\sim 30^\circ$ (ref. 11). Atl1 achieves greater DNA bending through synergistic amino terminus and binding site loop actions. The Atl1 N-terminal helix extends outward more than the corresponding AGT helix, which follows a loop leading towards the N-terminal domain (Figs 1a and 2a). This N-terminal extension pushes against the phosphate backbone of the complementary strand opposite the flipped nucleotide. Moreover, the binding site loop acts as a gate that switches between 'open' (Fig. 2b) and 'closed' (Fig. 2c) conformations of free and DNA-bound Atl1, respectively, with flanking glycines suggesting flexibility. This 'gating' action was proposed in AGT computational simulations²¹ but was not seen in crystal structures^{11,12}. The Atl1 binding site loop open-to-closed conformational switch appears suitable to have an active role in signaling by shifting covering Arg and Ile side chains to expose the C-terminal loop for possible intermolecular interactions.

If the closed, bent ATL–DNA complex is a platform for repair protein recognition, then we expect the complex to be stable; however, K_d was estimated by gel-shift assays to be only 0.41 μ M for TTHA1564 with O^6 -mG¹⁷. To test binding affinity, we measured Atl1 binding and

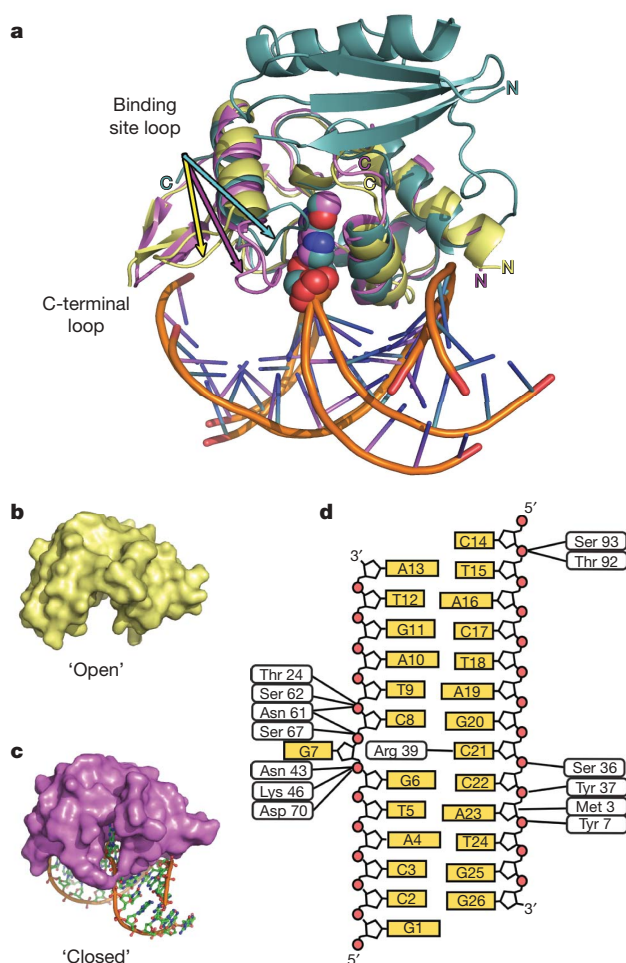


Figure 2 | Atf1 DNA binding and damage sculpting. **a**, Overlay of DNA-free (yellow) and DNA-bound (magenta) Atf1 with the human AGT (cyan) C-terminal domain. The rotated O^6 -mG (centre, spheres) is shown with the binding site loop that determines the open or closed conformation of Atf1. **b**, Atf1 molecular surface revealing an 'open' state. **c**, DNA-bound Atf1 molecular surface showing the protein 'closed' state. **d**, Atf1-DNA interaction schematic.

dissociation to and from oligonucleotides containing O^6 -mG, O^6 -pobG or abasic site by surface plasmon resonance (Fig. 3a). Atf1 binding to oligonucleotides containing alkyl-G was 0.02–0.3 nM, but binding to abasic site dsDNA was low and/or transient (Fig. 3a). Langmuir fit with mass transfer limitation indicates $k_{\text{ass}} = 1.21 \times 10^{-7} \pm 0.20 \times 10^{-7} \text{ M}^{-1} \text{ s}^{-1}$, $k_{\text{diss}} = 0.004 \pm 0.0006 \text{ s}^{-1}$ and $K_d = 0.35 \pm 0.04 \text{ nM}$ for O^6 -mG, and $k_{\text{ass}} = 2.20 \times 10^{-7} \pm 0.56 \times 10^{-7} \text{ M}^{-1} \text{ s}^{-1}$, $k_{\text{diss}} = 0.0003 \pm 0.00002 \text{ s}^{-1}$ and $K_d = 0.016 \pm 0.004 \text{ nM}$ for O^6 -pobG. Thus, whereas 'on' rates were similar for both lesions, the 'off' rate for the pob oligonucleotide was slower: higher affinity (K_d) for the pob oligonucleotide shows larger O^6 -alkyl groups are accommodated stably.

We also measured Atf1 binding to methylated double-stranded oligonucleotides by gel-shift assays (Fig. 3b and Supplementary Fig. 1 and Supplementary Table 3) and verified saturated complex stoichiometries by sedimentation equilibrium analysis (Fig. 3c and Supplementary Table 3). Dominant complexes for 13-base oligonucleotides in solution have 1:1 stoichiometry (Fig. 3b, c and Supplementary Table 3), consistent with our crystal structures, whereas 16-base oligonucleotides form 2:1 limiting complexes (Supplementary Fig. 1a and Supplementary Table 3) and 26-base oligonucleotides form 3:1 limiting complexes (Supplementary Fig. 1b and Supplementary Table 3). Saturated Atf1–nonmethylated DNA complex is formed without intermediate

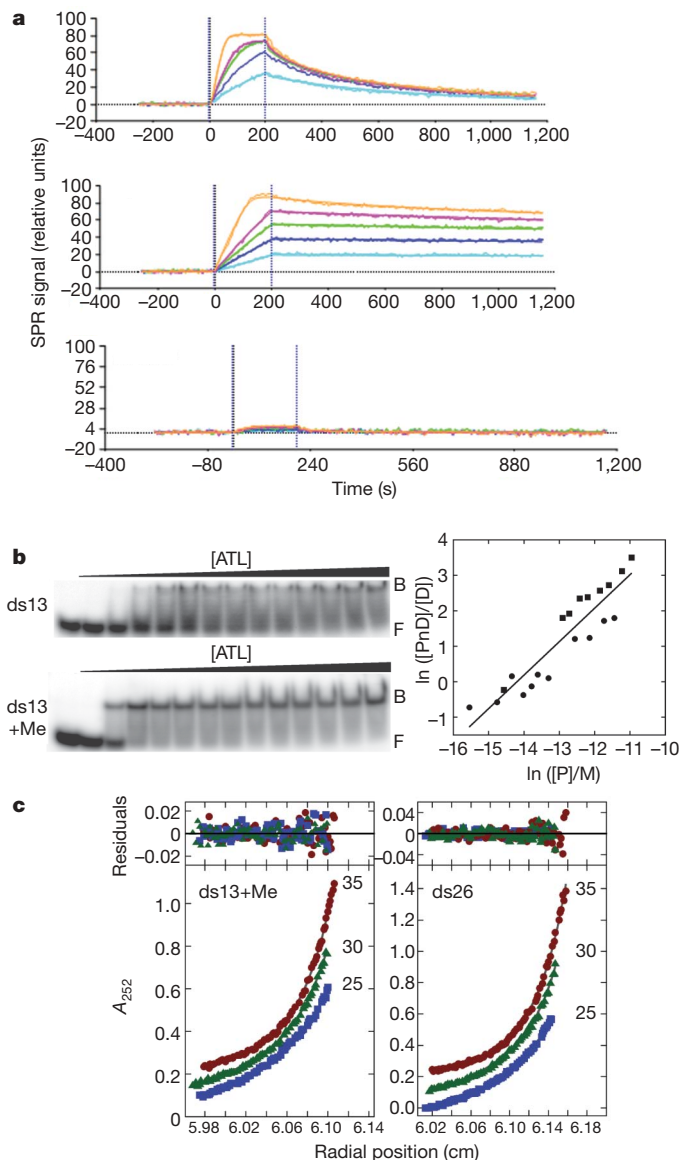


Figure 3 | Atf1 DNA lesion binding affinity and stoichiometry. **a**, Atf1 binding and dissociation for oligonucleotides containing O^6 -mG (top), O^6 -pobG (centre) or abasic site opposite cytosine (bottom). The coloured lines represent a series of Atf1 dilutions (0–4 nM for O^6 -mG and O^6 -pobG, 0–5 nM for abasic site). **b**, Gel-shift assays (left) for Atf1 binding normal (ds13) and O^6 -mG (ds13 + Me) 13-base oligonucleotide dsDNA; F and B denote DNA-free and -bound complexes. The associated O^6 -mG DNA binding isotherm analysis (right) shows two independent experiments (circles and squares), where P represents protein, D represents DNA and PnD represents protein–DNA complex containing n molecules of protein. Concentrations are given in molar units (M). (See Supplementary Information for further discussion.) **c**, Sedimentation equilibrium data for O^6 -mG 13-base oligonucleotide dsDNA complexes (left) and normal 26-base oligonucleotide dsDNA (right). Small, randomly distributed residuals (top panels) indicate models in which free protein, DNA and one protein–DNA complex equilibrate in solution. Calculated stoichiometries are 1.15 ± 0.08 for O^6 -mG 13-base oligonucleotide dsDNA and 3.03 ± 0.20 for normal 26-base oligonucleotide dsDNA.

accumulation, suggesting cooperative binding to nonmethylated DNA. DNAs containing O^6 -mG form 1:1 complexes before proceeding to saturation in an additional concerted step, suggesting that specific binding to O^6 -mG sites precedes build-up of cooperative assembly, consistent with the open-to-closed switch and a binding site size of ~ 8 bp. This differs from AGT's binding site size of 4 bp per protein²², possibly owing to AGT's added N-terminal domain and the open-to-closed switch (not seen in AGT) that exposes the C-terminal loop.

Connections of At1 to nucleotide excision repair

ATLs tightly bind oligonucleotides containing O^6 -alkylguanine and switch conformation to expose the C-terminal loop, indicating that ATL–DNA complex binding partners are possible, *in vivo*. In fact, the NER protein UvrA interacts with TTHA1564 (ref. 17). Similarly, far-western analysis reveals that *E. coli* Atl interacts with *E. coli* NER proteins UvrA and UvrC *in vitro* (Fig. 4a). *E. coli* DNA repair helicase IV (HelD) is also a potential *E. coli* Atl binding partner¹⁶. Interestingly, *S. pombe* Atl1 interacts with *E. coli* UvrA *in vitro* (Fig. 4b), indicating that ATLs conserve features across species for NER recognition.

To test for Atl1 functional genetic interactions with the NER pathway in fission yeast, we analysed *S. pombe* *atl1* and *rad13* (NER pathway XPG endonuclease human homologue that cuts 3' of DNA lesions²³) single and double deletants. We measured the ability of Atl1 to protect cells from *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG)-induced and spontaneous mutations (Fig. 4c–e). MNNG sensitivity of Δ *atl1* cells was complemented by a plasmid harbouring *atl1* (Supplementary Fig. 2), indicating that observed cellular phenotypes are due to *atl1* deletion. Atl1 inactivation causes approximately ninefold increased reversion rate of the *ade6-485* mutation, similar to the Δ *rad13* mutant (Fig. 4c, top). The MNNG-induced mutation rate is not further increased in the Δ *atl1* Δ *rad13* double mutant, revealing an epistatic relationship between Atl1 and Rad13 (Fig. 4c, top). These results are supported by spot tests (Fig. 4d) and clonogenic assays (Fig. 4e), indicating Atl1 and Rad13 are also epistatic for MNNG toxicity.

Strikingly, the increased spontaneous mutation rate of Δ *rad13* cells is suppressed to wild-type levels by additional Atl1 inactivation (Fig. 4c, bottom). This effect is not due to decreased cell survival, because all mutants tested here are viable (Supplementary Fig. 3). Clonogenic assays also revealed that Atl1 is epistatic with *S. pombe* Swi10 (Fig. 4f), but not Rhp14 (Fig. 4g) or Rad2 (Fig. 4h; homologues of human ERCC1, XPA and FEN1, respectively), for MNNG toxicity. The non-epistatic relationship between Atl1 and Rhp14 indicates that Rhp14 has an NER-independent function in response to MNNG, consistent with Rhp14 responses to other DNA-damaging agents (R.K. and O.F., unpublished data). Because Rad2 has a role in

long-patch base excision repair (BER)²⁴ and the alternative ultra-violet excision repair²⁵, lack of epistasis between *atl1* and *rad2* implies the two proteins work in different pathways. *Rad13* and *swi10* mutant phenotypes may reflect build-up of stable hard-to-repair ATL-complex intermediates in the absence of these NER proteins, indicating that ATL–DNA complexes may block alternative repair. Similarly, Atl1 protects *E. coli* cells against MNNG-induced alkylation damage (Supplementary Table 4). Thus, both microbial and eukaryotic genetic evidence indicates that ATL bridges DNA-alkylation base damage responses to NER.

Identification of novel ATLs

To see whether our structures may characterize other ATLs, we mapped sequence conservation of 197 ATL sequences based on the Atl1 structure (Supplementary Figs 4 and 5). For all ATLs, the most conserved residues line the lesion-binding pocket or act in DNA binding in our structures, suggesting that our Atl1 structures are paradigmatic for ATLs.

Significantly, our structure-based sequence analyses helped us to identify here the first ATL from any multicellular organism, the recently sequenced starlet sea anemone *Nematostella vectensis*²⁶ plus two archaeal ATLs from *Candidatus Korarchaeum cryptofilum*²⁷ and *Nanoarchaeum equitans*²⁸, ancestral to the two established phyla of archaea (GenBank accession numbers XM_001618690, YP_001736655 and NP_963633, respectively; Supplementary Figs 4 and 6). We verified that *N. vectensis* ATL blocks the alkyltransferase activity of human AGT (Supplementary Fig. 7), confirming it is an ATL. The *N. vectensis* genome, which aided in genome characterization of the long-extinct last common ancestor of all eumetazoans, is surprisingly more similar to vertebrates than to fruitflies or nematodes²⁶. Therefore, existence of ATL in this multicellular eukaryote, plus yeast, bacteria and ancestral archaea, shows that ATL is present in all three domains of life and argues that ATL was common to evolutionary branches before complex eukaryotes. This discovery suggests that complex eukaryotes and mammals will either have an ATL or have lost or replaced it with an analogous protein.

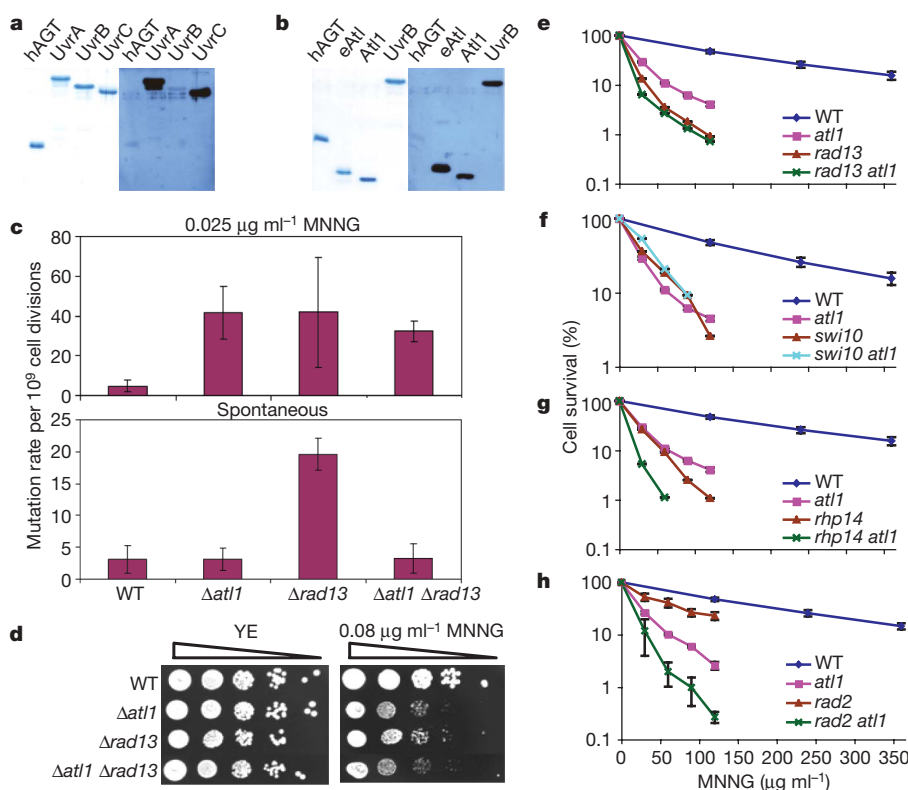


Figure 4 | Biochemical and genetic connection of Atl1 to NER. **a**, **b**, Coomassie-stained gel (left) and far-western blot (right) probed with Flag-tagged *E. coli* Atl (**a**) or Flag-UvrA (**b**). hAGT, human AGT; eAtl, *E. coli* Atl. **c**, MNNG-induced (top) and spontaneous (bottom) mutations of wild-type (WT), Δ *atl1*, Δ *rad13* and Δ *atl1* Δ *rad13* *S. pombe* strains. Results shown are mean \pm s.d.; $n \geq 3$. **d**, Atl1 and Rad13 are epistatic for MNNG toxicity. Serial dilutions of wild-type, Δ *atl1*, Δ *rad13* and Δ *atl1* Δ *rad13* *S. pombe* cells spotted on yeast extract (YE) plates or YE plates containing 0.08 μ g ml⁻¹ MNNG. **e**–**h**, Clonogenic assay, revealing that *atl1* is epistatic for MNNG toxicity with *rad13* (**e**) and *swi10* (**f**), but not *rhp14* (**g**) or *rad2* (**h**). Results shown are mean \pm s.e.m.; $n \geq 3$.

Discussion

Alkylated DNA base damage is classically repaired by direct damage reversal proteins or by lesion-specific DNA glycosylases, which excise modified bases to create abasic sites and initiate the BER pathway²⁹. These base repair processes differ from the versatile NER removal of bulky, unrelated, helix-distorting lesions by excising a lesion-containing DNA patch³⁰. Our combined results reveal a general mechanism for ATL to bind weakly distorting *O*⁶-alkylguanine lesions and recruit NER proteins (Fig. 5). We propose that ATL binding sculpts alkylguanine into a bulky lesion that is channelled into the NER pathway, explaining NER-mediated repair of *O*⁶-alkylguanine lesions^{31–33}.

ATL may be an unrecognized NER element, with analogues in many organisms, that targets endogenous alkylation damage to NER nucleases. In complex eukaryotes, the NER transcription-coupled repair (TCR) sub-pathway engages downstream damage recognition components of global genome repair (GGR), to effect lesion removal from the transcribed strand³⁴. The NER GGR sub-pathway is initiated by XPC recognition of bulky lesions and, like TCR, results in damage removal by incision on either side of the lesion³⁵. Because the *O*⁶-mG lesion is insufficient to block transcription³⁶, ATL binding may stall RNA polymerase to initiate TCR and/or promote lesion processing analogously to DDB2 of mammalian GGR. Atl1–DNA contacts are with the damaged strand, similar to DDB2 (ref. 37), consistent with possible undamaged strand binding by fission yeast XPC homologues Rhp41 or Rhp42, as shown for *S. cerevisiae* XPC orthologue Rad4 (ref. 38).

ATLs are not alkyltransferases^{14,15,17} or glycosylases^{14,15}, but inhibit AGT^{14,15} (Supplementary Fig. 7). Lack of epistasis for MNNG-induced cell killing between *atl1* and *rad2* shows that ATL is not a

long-patch BER or alternative ultraviolet excision repair protein. Yet, ATL damage recognition resembles that of AGT and BER glycosylases rather than that of NER proteins. It uses a positively charged channel for lesion-binding and 180° nucleotide flipping, which allow protein handoffs without release of toxic and mutagenic DNA intermediates, a hallmark of BER and recombination repair pathways^{39–43}.

ATL binding targets base damage to NER, showing how proteins that bind damage, but do not repair it, may redirect lesion processing⁴⁴. First, ATL binds base damage analogously to AGT and BER glycosylases^{2,45}, but presents damage to NER similarly to DDB2. Second, in some organisms ATL can block AGT *O*⁶-alkylguanine damage recognition and redirect base repair to NER, constituting a crosstalk pathway connection⁴⁴, as proposed for AGT^{33,46} and recently for *E. coli* Atl²⁰. Third, ATL redirects endogenous damage from other repair pathways to NER, because the $\Delta rad13$ mutator is rescued in $\Delta atl1 \Delta rad13$ mutants. Fourth, ancestral archaeal ATLs are ATL–Endo V fusions, suggesting that ATL and Endo V act together in a coordinated pathway⁴⁷, with the BER nuclease Endo V serving a possible XPG-like function in these organisms, as AGT–Endo V fusion proteins retain both activities⁴⁸. By the Rosetta Stone evolution hypothesis for protein interactions⁴⁷, ATL–Endo V fusions imply that ATL provides a primordial connection joining BER and NER. Indeed, recent structures of Endo V (ref. 49) and NER complex DDB1–DDB2 (ref. 37) support such an ancient BER–NER connection by revealing a mutual, wedge-based binding mechanism⁵⁰. Thus, non-enzymatic nucleotide flipping emerges as a surprisingly general mechanism to channel specific base damage into the general damage NER pathway by handoff from a non-enzymatic complex.

METHODS SUMMARY

Atl1 purification, crystallization, X-ray diffraction data collection and structural refinement. C-terminally 6× His-tagged Atl1 was expressed in JM109 cells and purified over Ni-NTA agarose and Superdex 75 columns. Atl1–*O*⁶-mG- and Atl1–*O*⁶-pobG DNA complexes were prepared at a 1.5:1 DNA:protein molar ratio. Crystals were grown by hanging-drop vapour diffusion. Diffraction data were collected at ALS beamline 12.3.1 for Atl1 and at SSRL beamline 11-1 for Atl1–*O*⁶-mG and Atl1–*O*⁶-pobG DNA complexes, and were processed with HKL2000. Structures were solved by molecular replacement with Phaser, using a modified wild-type Ada-C (PDB accession code 1SFE) as a search model for Atl1, and the refined Atl1 structure as a search model for Atl1–DNA complexes. Crystallographic refinement was done with Crystallography & NMR System (CNS), and Xfit was used for manual model building.

DNA binding by Atl1. Oligonucleotide–Atl1 interactions were analysed by an electrophoretic mobility shift assay using standard methods and by surface plasmon resonance with biotinylated *O*⁶-mG-, *O*⁶-pobG- or AP-site-containing or control oligonucleotides immobilized on a streptavidin-coated surface of a BioRad NLC chip and serial dilutions of Atl1 applied to the chip. DNA complex stoichiometries formed under protein saturation conditions were established by sedimentation equilibrium analysis.

Other biochemical assays. AGT inhibition assays were performed as described previously. Far western analyses were performed using standard methods.

Atl1 expression in *S. pombe*. *S. pombe* strains originated from GM4 (*h⁺ atl1::ura4 ura4-D18 leu1-32 his7-366 ade6-M210*) or RO131 (*h⁺ rad13::kanMX ura4-D18 his3-D1*). MNNG sensitivity was determined by agar plate and clonogenic assays. Mutation rates were determined as reversions of *ade6-485* to *Ade⁺*.

Atl1 expression in *E. coli*. The pQE-30 empty vector was expressed in *E. coli* GWR109 *ada⁺ ogt⁺* and pQE-30 or pQE-Atl1 in *E. coli* GWR109 *ada⁺ ogt⁺ atl⁺*. Mutation frequencies were determined as the number of MNNG-induced Rifampicin-resistant mutants (Rif^R) per 10⁸ surviving cells.

Received 28 January; accepted 17 April 2009.

- Pegg, A. E. Repair of *O*⁶-alkylguanine by alkyltransferases. *Mutat. Res.* **462**, 83–100 (2000).
- Tubbs, J. L., Pegg, A. E. & Tainer, J. A. DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by *O*⁶-alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. *DNA Repair (Amst.)* **6**, 1100–1115 (2007).
- Loechler, E. L., Green, C. L. & Essigmann, J. M. *In vivo* mutagenesis by *O*⁶-methylguanine built into a unique site in a viral genome. *Proc. Natl Acad. Sci. USA* **81**, 6271–6275 (1984).

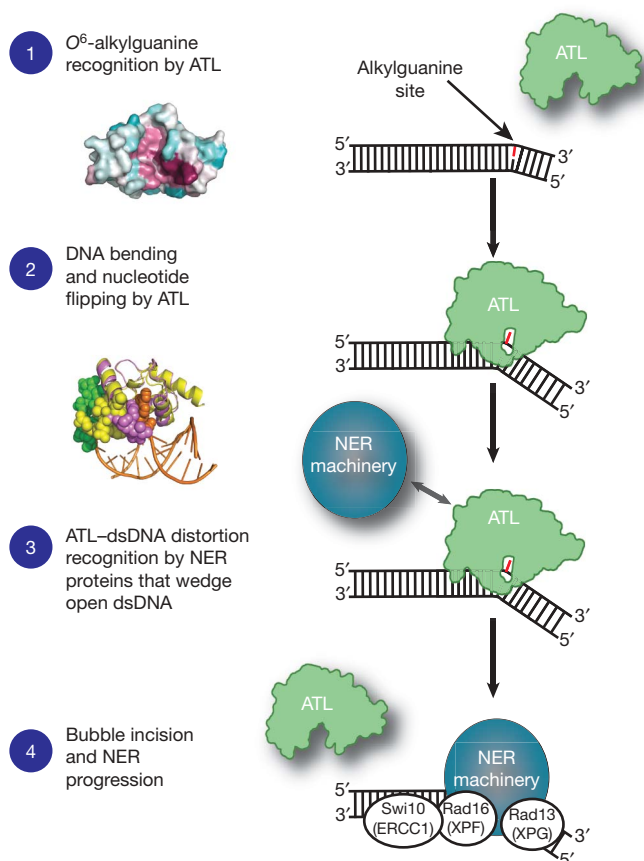


Figure 5 | Alkyl-G lesion recognition allows NER repair of relatively non-distorting base lesions. The distorted, stable ATL–DNA complex creates a platform to recruit NER enzymes to *O*⁶-alkyl-G lesions. This general model is based on our combined structural, biochemical and genetic results.

4. Pauly, G. T., Hughes, S. H. & Moschel, R. C. Comparison of mutagenesis by O⁶-methyl- and O⁶-ethylguanine and O⁴-methylthymine in *Escherichia coli* using double-stranded and gapped plasmids. *Carcinogenesis* **19**, 457–461 (1998).
5. Margison, G. P. & Santibáñez-Koref, M. F. O⁶-Alkylguanine-DNA alkyltransferase: role in carcinogenesis and chemotherapy. *Bioessays* **24**, 255–266 (2002).
6. Mitra, S. & Kaina, B. Regulation of repair of alkylation damage in mammalian genomes. *Prog. Nucleic Acid Res. Mol. Biol.* **44**, 109–142 (1993).
7. Pegg, A. E., Dolan, M. E. & Moschel, R. C. Structure, function and inhibition of O⁶-alkylguanine-DNA alkyltransferase. *Prog. Nucleic Acid Res. Mol. Biol.* **51**, 167–223 (1995).
8. Daniels, D. S. & Tainer, J. A. Conserved structural motifs governing the stoichiometric repair of alkylated DNA by O⁶-alkylguanine-DNA alkyltransferase. *Mutat. Res.* **460**, 151–163 (2000).
9. Daniels, D. S. *et al.* Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding. *EMBO J.* **19**, 1719–1730 (2000).
10. Wibley, J. E. A., Pegg, A. E. & Moody, P. C. E. Crystal structure of the human O⁶-alkylguanine-DNA alkyltransferase. *Nucleic Acids Res.* **28**, 393–401 (2000).
11. Daniels, D. S. *et al.* DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nature Struct. Mol. Biol.* **11**, 714–720 (2004).
12. Duguid, E. M., Rice, P. A. & He, C. The structure of the human AGT protein bound to DNA and its implications for damage detection. *J. Mol. Biol.* **350**, 657–666 (2005).
13. Margison, G. P. *et al.* Alkyltransferase-like proteins. *DNA Repair (Amst.)* **6**, 1222–1228 (2007).
14. Pearson, S. J., Ferguson, J., Santibáñez-Koref, M. & Margison, G. P. Inhibition of O⁶-methylguanine-DNA methyltransferase by an alkyltransferase-like protein from *Escherichia coli*. *Nucleic Acids Res.* **33**, 3837–3844 (2005).
15. Pearson, S. J. *et al.* A novel DNA damage recognition protein in *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **34**, 2347–2354 (2006).
16. Chen, C. S. *et al.* A proteome chip approach reveals new DNA damage recognition activities in *Escherichia coli*. *Nature Methods* **5**, 69–74 (2008).
17. Morita, R., Nakagawa, N., Kuramitsu, S. & Masui, R. An O⁶-methylguanine-DNA methyltransferase-like protein from *Thermus thermophilus* interacts with a nucleotide excision repair protein. *J. Biochem.* **144**, 267–277 (2008).
18. Wang, L. *et al.* Pyridyloxobutyl adduct O⁶-[4-oxo-4-(3-pyridyl)butyl]guanine is present in 4-(acetoxymethylnitrosamino)-1-(3-pyridyl)-1-butanone-treated DNA and is a substrate for O⁶-alkylguanine-DNA alkyltransferase. *Chem. Res. Toxicol.* **10**, 562–567 (1997).
19. Mijal, R. S. *et al.* The repair of the tobacco specific nitrosamine derived adduct O⁶-[4-Oxo-4-(3-pyridyl)butyl]guanine by O⁶-alkylguanine-DNA alkyltransferase variants. *Chem. Res. Toxicol.* **17**, 424–434 (2004).
20. Mazon, G. *et al.* The alkyltransferase-like *ybaZ* gene product enhances nucleotide excision repair of O⁶-alkylguanine adducts in *E. coli*. *DNA Repair (Amst.)*. (in the press).
21. Hu, J., Ma, A. & Dinner, A. R. A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT. *Proc. Natl Acad. Sci. USA* **105**, 4615–4620 (2008).
22. Rasimas, J. J., Pegg, A. E. & Fried, M. G. DNA-binding mechanism of O⁶-alkylguanine-DNA alkyltransferase. Effects of protein and DNA alkylation on complex stability. *J. Biol. Chem.* **278**, 7973–7980 (2003).
23. O'Donovan, A. *et al.* XPG endonuclease makes the 3' incision in human DNA nucleotide excision-repair. *Nature* **371**, 432–435 (1994).
24. Klungland, A. & Lindahl, T. Second pathway for completion of human DNA base excision-repair: Reconstitution with purified proteins and requirement for Dnase IV (FEN1). *EMBO J.* **16**, 3341–3348 (1997).
25. Yonemasu, R. *et al.* Characterization of the alternative excision repair pathway of UV-damaged DNA in *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **25**, 1553–1558 (1997).
26. Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).
27. Elkins, J. G. *et al.* A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl Acad. Sci. USA* **105**, 8102–8107 (2008).
28. Waters, E. *et al.* The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl Acad. Sci. USA* **100**, 12984–12988 (2003).
29. Sedgwick, B. Repairing DNA-methylation damage. *Nature Rev. Mol. Cell Biol.* **5**, 148–157 (2004).
30. Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nature Rev. Mol. Cell Biol.* **9**, 958–970 (2008).
31. Samson, L., Thomale, J. & Rajewsky, M. F. Alternative pathways for the *in vivo* repair of O⁶-alkylguanine and O⁴-alkylthymine in *Escherichia coli*: the adaptive response and nucleotide excision repair. *EMBO J.* **7**, 2261–2267 (1988).
32. Voigt, J. M., Van Houten, B., Sancar, A. & Topal, M. D. Repair of O⁶-methylguanine by ABC excinuclease of *Escherichia coli* *in vitro*. *J. Biol. Chem.* **264**, 5172–5176 (1989).
33. Edara, S., Kanugula, S. & Pegg, A. E. Expression of the inactive C145A mutant human O⁶-alkylguanine-DNA alkyltransferase in *E. coli* increases cell killing and mutations by N-methyl-N'-nitro-N-nitrosoguanidine. *Carcinogenesis* **20**, 103–108 (1999).
34. Mellon, I. Transcription-coupled repair: a complex affair. *Mutat. Res.* **577**, 155–161 (2005).
35. Branum, M. E., Reardon, J. T. & Sancar, A. DNA repair excision nuclease attacks undamaged DNA. A potential source of spontaneous mutations. *J. Biol. Chem.* **276**, 25421–25426 (2001).
36. Viswanathan, A. & Doetsch, P. W. Effects of nonbulky DNA base damages on *Escherichia coli* RNA polymerase-mediated elongation and promoter clearance. *J. Biol. Chem.* **273**, 21276–21281 (1998).
37. Scrima, A. *et al.* Structural basis of UV DNA-damage recognition by the DDB1–DDB2 complex. *Cell* **135**, 1213–1223 (2008).
38. Min, J. H. & Pavletich, N. P. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **449**, 570–575 (2007).
39. Mol, C. D., Izumi, T., Mitra, S. & Tainer, J. A. DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination. *Nature* **403**, 451–456 (2000).
40. Chapados, B. R. *et al.* Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell* **116**, 39–50 (2004).
41. Parikh, S. S. *et al.* Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc. Natl Acad. Sci. USA* **97**, 5083–5088 (2000).
42. Garcin, E. D. *et al.* DNA apurinic-apyrimidinic site binding and excision by endonuclease IV. *Nature Struct. Mol. Biol.* **15**, 515–522 (2008).
43. Williams, R. S. *et al.* Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. *Cell* **135**, 97–109 (2008).
44. Cline, S. D. & Hanawalt, P. C. Who's on first in the cellular response to DNA damage? *Nature Rev. Mol. Cell Biol.* **4**, 361–372 (2003).
45. Hitomi, K., Iwai, S. & Tainer, J. A. The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA Repair (Amst.)* **6**, 410–428 (2007).
46. Hickman, M. J. & Samson, L. D. Role of DNA mismatch repair and p53 in signaling induction of apoptosis by alkylating agents. *Proc. Natl Acad. Sci. USA* **96**, 10764–10769 (1999).
47. Marcotte, E. M. *et al.* Detecting protein function and protein–protein interactions from genome sequences. *Science* **285**, 751–753 (1999).
48. Kanugula, S., Pauly, G. T., Moschel, R. C. & Pegg, A. E. A bifunctional DNA repair protein from *Ferroplasma acidimanus* exhibits O⁶-alkylguanine-DNA alkyltransferase and endonuclease V activities. *Proc. Natl Acad. Sci. USA* **102**, 3617–3622 (2005).
49. Dalhus, B. *et al.* Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nature Struct. Mol. Biol.* **16**, 138–143 (2009).
50. Schärer, O. D. & Campbell, A. J. Wedging out DNA damage. *Nature Struct. Mol. Biol.* **16**, 102–104 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. C. Vu and J. Gong for aiding in the synthesis of O⁶-pobG oligomers, M. N. Boddy, J. Prudden and A. Sarker for performing genetics and biochemical experiments, G. Guenther, S. Pebernard, R. S. Williams, J. J. Perry, B. R. Chapados, M. Björås, D. S. Shin, K. Hitomi, C. Hitomi, E. D. Getzoff, G. Williams, S. Tsutakawa and P. K. Cooper for suggestions, and the staff at the Advanced Light Source (ALS) SIBYLS beamline and the Stanford Synchrotron Radiation Laboratory (SSRL). Operations at SSRL and ALS are supported by the US Department of Energy and NIH. This work was supported by National Institutes of Health grants CA097209 (J.A.T., A.E.P.), CA018137 (A.E.P.), GM070662 (M.G.F.), and CA59887 (L.A.P.), The Skaggs Institute for Chemical Biology (J.L.T.), North West Cancer Research Fund grant CR675 (O.F.), Cancer Research-UK (G.P.M.) and CHEMORES (G.P.M.).

Author Contributions M.D.K. and J.L.T. purified AtI1 protein and prepared AtI1-oligomer complexes for crystallization. A.S.A. and J.L.T. crystallized AtI1 and collected, processed and refined X-ray data. O⁶-pobG oligomers were synthesized by L.A.P. for crystallization and by C.M. and D.M.W. for surface plasmon resonance. A.J.W. and B.V. designed and synthesized oligonucleotides that contributed to the surface plasmon resonance data. A.M. and A.J.W. produced and characterized pure AtI1 protein for the surface plasmon resonance analyses. G.M. and M.T. performed surface plasmon resonance analyses. M.M. and M.G.F. performed electrophoretic mobility shift assays and analytical ultracentrifugation experiments and analysed the results. V.L. and A.B. generated AtI1 single and double deletants and AtI1-complement in *S. pombe* and carried out spot and clonogenic assays. R.K. and O.F. carried out the mutation assays in *S. pombe*. S.K. prepared constructs for and purified AtI1, *E. coli* AtI, *N. vectensis* ATL, AGT C145S, UvrA, UvrB and UvrC, and performed far western analyses, AtI1 expression assays in *E. coli* and ATL inhibition assays. M.F.S.-K. contributed intellectually to the initiation and design of the studies at the Paterson Institute. O.F., G.P.M., A.E.P. and J.A.T. provided intellectual guidance and research support. J.L.T. and J.A.T. wrote the paper. All authors discussed the results and manuscript.

Author Information Atomic coordinates and structure factors for the reported crystal structures have been deposited in the Protein Data Bank under accession codes 3GVA (AtI1), 3GX4 (AtI1–O⁶-mG-DNA) and 3GYH (AtI1–O⁶-pobG-DNA). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.A.T. (jat@scripps.edu) or A.E.P. (aep1@psu.edu).

LETTERS

Earth's transmission spectrum from lunar eclipse observations

Enric Pallé¹, María Rosa Zapatero Osorio¹, Rafael Barrena¹, Pilar Montañés-Rodríguez¹ & Eduardo L. Martín^{1,2}

Of the 342 planets so far discovered¹ orbiting other stars, 58 'transit' the stellar disk, meaning that they can be detected through a periodic decrease in the flux of starlight². The light from the star passes through the atmosphere of the planet, and in a few cases the basic atmospheric composition of the planet can be estimated^{3–5}. As we get closer to finding analogues of Earth^{6–8}, an important consideration for the characterization of extrasolar planetary atmospheres is what the transmission spectrum of our planet looks like. Here we report the optical and near-infrared transmission spectrum of the Earth, obtained during a lunar eclipse. Some biologically relevant atmospheric features that are weak in the reflection spectrum⁹ (such as ozone, molecular oxygen, water, carbon dioxide and methane) are much stronger in the transmission spectrum, and indeed stronger than predicted by modelling^{10,11}. We also find the 'fingerprints' of the Earth's ionosphere and of the major atmospheric constituent, molecular nitrogen (N₂), which are missing in the reflection spectrum.

The characterization of spectral features in our planet's transmission spectrum can be achieved through observations of the light reflected from the Moon towards the Earth during a lunar eclipse, which resembles the observing geometry during a planetary transit. At that time, the reflected sunlight from the lunar surface within the Earth's umbra will be entirely dominated by the fraction of sunlight that is transmitted through an atmospheric ring located along the Earth's day–night terminator (Supplementary Fig. 1). Observations of the lunar eclipse on 16 August 2008 have allowed us to characterize the Earth's spectrum as if it were observed from an astronomical distance during a transit in front of the Sun. Except for some early attempts^{12,13} with photographic plates, spectroscopic lunar eclipse observations in visible or near-infrared wavelengths have not previously been undertaken.

The Earth's transmission spectrum can be calculated from the brightness ratio of the light reflected by the lunar surface when in the umbra, in the penumbra and out of the eclipse (Supplementary Information). The resulting transmission spectrum is shown in Fig. 1a, where simultaneous optical and near-infrared observations with the William Herschel and Nordic Optical Telescopes are inter-calibrated to provide continuous wavelength coverage from 0.36 to 2.40 μm . It is known that the Earth's optical and near-infrared transmission spectrum is red¹⁴—that is, more solar flux successfully passes through the atmosphere at longer wavelengths, as can be inferred from simple naked-eye observations of a lunar eclipse, or of a sunset/sunrise. The rising nature of the transmission spectrum continuum towards longer wavelengths (possibly the most remarkable feature of the data shown in Fig. 1a) is caused by the Rayleigh scattering of air, which, in addition to the ozone Chappuis band absorption between 0.375 and 0.650 μm , is rather efficient in scavenging short wavelength radiation through a long atmospheric path.

During a transit, the starlight that passes through the planet's atmosphere travels through a much larger atmospheric path than the starlight that is directly reflected by the planet. This causes the reflection and transmission spectra to be different (see Fig. 1b). In transmission, the Earth is brighter in the near-infrared range, particularly at 2.2 μm .

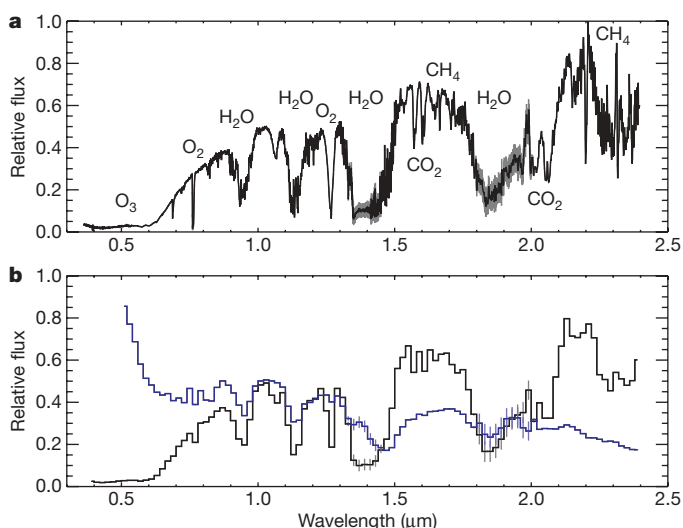


Figure 1 | Earth's visible and near-infrared transmission and reflection spectra. The Earth's transmission spectrum is a proxy for Earth observations during a primary transit as seen beyond the Solar System, while the reflection spectrum is a proxy for the observations of Earth as an exoplanet by direct observation after removal of the Sun's spectral features. **a**, The transmission spectrum, with some of the major atmospheric constituents marked. The spectrum has a resolution of 0.00068 μm in the optical (resolving power, $R \approx 960$) and 0.0013–0.0024 μm in the near-infrared ($R \approx 920$). A detailed atlas of the transmission spectrum with identifications of the main atomic and molecular absorption features is available (Supplementary Fig. 3). **b**, A comparison between the Earth's transmission (black) and reflection (blue) spectra. Both spectra have been degraded to a spectral resolution of 0.02 μm and normalized at the same flux value at around 1.2 μm . It is readily seen from the figure that the reflection spectrum shows increased Rayleigh reflectance in the blue. It is also noticeable how most of the molecular spectral bands are weaker, and some non-existent, in the reflection spectrum. In **a** and **b**, the noise (r.m.s.) of the spectra, which takes into account the corrections for the strongest local telluric features, is plotted point-per-point (in grey for black lines, in light blue for dark blue line) along with the spectra, although, for most spectral regions, the size of the error bars is comparable to the width of the line. The quality of the transmission data are measured in terms of signal-to-noise ratio, which goes from ~ 100 (at the deepest absorptions features and the blue optical wavelengths) up to ~ 400 (at the largest values of the relative fluxes).

¹Instituto de Astrofísica de Canarias, Vía Láctea s/n, E38205 La Laguna, Tenerife, Spain. ²Physics Department, University of Central Florida, PO Box 162385, Orlando, Florida 32816, USA.

Table 1 | Equivalent widths of the major molecular absorption features in Earth's spectra

Atmospheric species	Wavelength interval (μm)	Equivalent width, transmission (\AA)	Equivalent width, reflection (\AA)
O ₂ -a	0.6858–0.6952	17.0 \pm 0.4	13.5 \pm 1.0
O ₂ -b	0.7570–0.7706	55.0 \pm 0.3	47.8 \pm 1.0
H ₂ O-a	0.7133–0.7342	12.2 \pm 1.0	34.0 \pm 1.0
H ₂ O-b	0.8057–0.8400	21.2 \pm 1.0	50.8 \pm 5.0
H ₂ O-c	0.8884–0.9966	331.1 \pm 9.0	205.3 \pm 6.0
H ₂ O-d	1.0870–1.1755	381.0 \pm 9.0	160.71 \pm 10.0
H ₂ O-e	1.3000–1.5212	1,300.0 \pm 30.0	623.4 \pm 30.0
H ₂ O-f	1.7586–1.9824	1,111.0 \pm 30.0	350.0 \pm 80.0
O ₂ •O ₂ -a	0.5556–0.6200	71.1 \pm 4.0	\leq 50.0
O ₂ •O ₂ -b	1.0295–1.0872	69.0 \pm 2.0	<5.0
O ₂ O ₂ •O ₂ O ₂ •N ₂ -ab	1.2347–1.2853	178.9 \pm 9.0	24.2 \pm 3.0
CO ₂ -a	1.5629–1.5908	52.6 \pm 1.0	10.0 \pm 3.0
CO ₂ -b	1.5908–1.6214	54.8 \pm 1.0	12.0 \pm 3.0
CO ₂ -c	1.9906–2.0373	150.6 \pm 1.0	23.0 \pm 5.0
CO ₂ -d	2.0373–2.0789	127.6 \pm 1.0	<15.0
CH ₄ -a	1.6214–1.6750	58.0 \pm 1.0	<15.0
CH ₄ -bcd	2.1390–2.4000	>790	>280.0

Equivalent widths are measured with respect to the pseudo-continuum that is absorbed by much wider features, like Rayleigh scattering or extended water bands. The second column is indicative of the wavelength range over which the various features are integrated. In all cases, except for a few features in the optical, the equivalent widths are larger in the transmission spectrum than in the reflectance data. The indices -a to -f in the first column are used to compare with the filter sets defined in Supplementary Table 1.

In contrast, the Earth's reflection spectrum appears blue¹⁵ because the very same Rayleigh scattering effect expels the short wavelength radiation back to space. Because of the blue colour, when observed from an astronomical distance, the Earth is often referred to as the pale blue dot, but in transmission, the pale blue dot becomes the pale red dot.

Moreover, the transmission spectrum (Supplementary Fig. 3) presents strong absorption features produced by molecular oxygen and oxygen collision complexes, including collisions between O₂ and N₂, which are significantly more intense than in the reflection spectrum. Oxygen collision complexes are van der Waals molecules, also known as dimers^{16–18}, which can be used in combination with other molecular oxygen bands to derive an averaged atmospheric column density of N₂. This is important, because although N₂ is the major atmospheric component (78.08% by volume), it lacks any marked electronic transition. The strength of these bands in the transmission spectrum implies that atmospheric dimers may become a major subject of study for the interpretation of rocky exoplanet transmission spectra and their atmospheric characterization. The atmospheric spectral bands of O₃, O₂, H₂O, CO₂ and CH₄ are readily distinguishable in the transmission spectrum (Supplementary Fig. 3). Trace amounts of N₂O, OClO and NO₂ (a gas mainly produced by human activities) might also be present in our data, but their detection will need future detailed modelling efforts to fit the observations.

The presence of the Earth's ionosphere is also revealed in our transmission spectrum through the detection of relatively weak and narrow absorption lines corresponding to singly ionized calcium atoms (Ca II, Supplementary Fig. 3). Calcium is the sixth most abundant element on Earth. It is possible that other, more abundant^{19,20} ionospheric species, such as singly ionized magnesium (Mg II), could be detectable at shorter wavelengths that are not covered by our data. The neutral atomic resonance doublet of sodium (Na I) is embedded in a quite strong and broad absorption due to dimers, ozone and Rayleigh scattering, and only the core of the doublet is detected at 0.5898 μm .

In the coming years, space missions such as the James Webb Space Telescope will perhaps have the opportunity to perform transit spectroscopy of rocky exoplanets^{21,22}. The faint signal of the planet atmosphere will be mixed with the light of the parent star, and it is foreseen that many transits will have to be observed²³ before a good quality transmission spectrum of an exo-Earth can be obtained. Our empirical transmission spectrum suggests, however, that retrieving the major planetary signals might be easier than model calculations suggest. Observations of about 20–30 one-hour transits should yield the detection of the major spectral features in the transmission spectrum of an Earth-like planet around a low-mass M-type star in the solar neighbourhood. With this goal in mind, it is necessary to determine and quantify the observational requirements to detect each specific

molecule using the transmission spectrum presented here. It is also useful to compare the potential results from two exoplanet atmospheric characterization techniques: direct detection and transit spectroscopy.

Observations of the earthshine—the light reflected from the dark side of the Moon—are often obtained and studied as a proxy for direct extrasolar planet observations^{9,24–27}, mainly using chronographic or interferometric techniques^{28,29}. Here, we were able to carry out earthshine observations with the same instrumental set-up as that during our lunar eclipse observations (see Supplementary Information for details about our data acquisition and analysis). The data are displayed in Fig. 1b for comparison with the transmission spectrum.

The depth of the features in the reflection spectrum of the Earth can vary depending on (1) changes in the properties of the surface from which part of the light is being reflected, and (2) changes in the cloud coverage of the globe^{25,30}. However, these variations are typically small (\leq 6%) and they do not modify significantly the shape of the data. In the transmission spectrum, there is a negligible contribution from light reflecting from the Earth's surface, thus only the changes in cloudiness can affect the depth of the absorption features, and we expect the amplitude of the spectral variability to be at most of the same order as in the reflection spectrum.

In Table 1 are listed the major molecular features in the transmission and reflection spectra, together with their equivalent widths measured over the original data with a resolution of 0.0013–0.0024 μm . Even at very low signal-to-noise ratios, the major atmospheric components remain marginally detectable in the transmission spectrum, but not in the reflection spectrum (see Supplementary Information for a detailed analysis of confidence of detection of atmospheric features at different signal-to-noise levels). Thus, the transmission spectrum can provide much more information about the atmospheric composition of a rocky planet than the reflection spectrum can.

Received 30 December 2008; accepted 8 April 2009.

1. The Extrasolar Planets Encyclopaedia. (<http://exoplanet.eu>) (accessed 18 March 2009).
2. Seager, S. & Sasselov, D. D. Theoretical transmission spectra during extrasolar giant planet transits. *Astrophys. J.* **537**, 916–921 (2000).
3. Charbonneau, D., Brown, T. M., Noyes, R. W. & Gilliland, R. L. Detection of an extrasolar planet atmosphere. *Astrophys. J.* **568**, 377–384 (2002).
4. Tinetti, G. *et al.* Water vapour in the atmosphere of a transiting extrasolar planet. *Nature* **448**, 169–171 (2007).
5. Swain, M. R., Vasisth, G. & Tinetti, G. The presence of methane in the atmosphere of an extrasolar planet. *Nature* **452**, 329–331 (2008).
6. Mayor, M. *et al.* The HARPS search for southern extra-solar planets. XIII. A planetary system with 3 super-Earths (4.2, 6.9, and 9.2 M_J). *Astron. Astrophys.* **493**, 639–644 (2009).

7. Borde, P., Rouan, D. & Leger, A. Exoplanet detection capability of the COROT space mission. *Adv. Space Res.* **405**, 1137–1144 (2003).
8. Basri, G., Borucki, W. J. & Koch, D. The Kepler Mission: A wide-field transit search for terrestrial planets. *N. Astron. Rev.* **49**, 478–485 (2005).
9. Montañés-Rodríguez, P., Palle, E., Goode, P. R., Hickey, J. & Koonin, S. E. Globally integrated measurements of the Earth's visible spectral albedos. *Astrophys. J.* **629**, 1175–1182 (2005).
10. Ehrenreich, D., Tinetti, G., Lecavelier Des Etangs, A., Vidal-Madjar, A. & Selsis, F. The transmission spectrum of Earth-size transiting planets. *Astron. Astrophys.* **448**, 379–393 (2006).
11. Miller-Ricci, E., Seager, S. & Sasselov, D. The atmospheric signatures of super-Earths: How to distinguish between hydrogen-rich and hydrogen-poor atmospheres. *Astrophys. J.* **690**, 1056–1067 (2009).
12. Slipher, V. M. On the spectrum of the eclipsed Moon. *Astron. Nachr.* **199**, 103 (1914).
13. Moore, J. H. & Brigham, L. A. The spectrum of the eclipsed Moon. *Publ. Astron. Soc. Pacif.* **39**, 223–226 (1927).
14. Shane, C. D. Photographs of the lunar eclipse of June. *Publ. Astron. Soc. Pacif.* **39**, 226–228 (1927).
15. Sagan, C., Thompson, W. R., Carlson, R., Gurnett, D. & Hord, C. A search for life on Earth from the Galileo spacecraft. *Nature* **365**, 715–718 (1993).
16. Calo, J. & Narcisi, R. Van der Waals molecules — possible roles in the atmosphere. *Geophys. Res. Lett.* **7**, 289–292 (1980).
17. Klemperer, W. & Vaida, V. Molecular complexes in close and far away. *Proc. Natl Acad. Sci. USA* **103**, 10584–10588 (2006).
18. Solomon, S., Portmann, R. W., Sanders, R. W. & Daniel, J. S. Absorption of solar radiation by water vapor, oxygen, and related collision pairs in the Earth's atmosphere. *J. Geophys. Res.* **103** (D4), 3847–3858 (1998).
19. Zbinden, P. A., Hidalgo, M. A., Eberhardt, P. & Geiss, J. Mass spectrometer measurements of the positive ion composition in the D- and E-regions of the ionosphere. *Planet. Space Sci.* **23**, 1621–1642 (1975).
20. Kopp, E. On the abundance of metal ions in the lower ionosphere. *J. Geophys. Res.* **102** (A5), 9667–9674 (1997).
21. Gardner, J. P. et al. The James Webb Space Telescope. *Space Sci. Rev.* **123**, 485–606 (2006).
22. Charbonneau, D. & Deming, D. The dynamics-based approach to studying terrestrial exoplanets. Preprint at <http://arXiv.org/abs/0706.1047> (2007).
23. Kaltenegger, L. & Traub, W. A. Transits of Earth-like planets. *Astrophys. J.* (in the press); preprint at <http://arXiv.org/abs/0903.3371v2> (2009).
24. Woolf, N. J., Smith, P. S., Traub, W. A. & Jucks, K. W. The spectrum of earthshine: A pale blue dot observed from the ground. *Astrophys. J.* **574**, 430–433 (2002).
25. Montañés-Rodríguez, P. P., Palle, E. & Goode, P. R. Vegetation signature in the observed globally-integrated spectrum of Earth: modeling the red edge strength using simultaneous cloud data and applications for extrasolar planets. *Astrophys. J.* **651**, 544–552 (2006).
26. Turnbull, M. C. et al. Spectrum of a habitable world: Earthshine in the near-infrared. *Astrophys. J.* **644**, 551–559 (2006).
27. Palle, E., Goode, P. R., Montañés-Rodríguez, R. & Koonin, S. E. Changes in the Earth's reflectance over the past two decades. *Science* **304**, 1299–1301 (2004).
28. Cash, W. Detection of Earth-like planets around nearby stars using a petal-shaped occulter. *Nature* **442**, 51–53 (2006).
29. Lindensmith, C. Terrestrial planet finder: technology development plans. Optical, infrared, and millimeter space telescopes. *Proc. SPIE* **5487**, 1226–1233 (2004).
30. Palle, E., Ford, E. B., Seager, S., Montañés-Rodríguez, P. & Vázquez, M. Identifying the rotation rate and the presence of dynamic weather on extrasolar Earth-like planets from photometric observations. *Astrophys. J.* **676**, 1319–1329 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Grundahl and J. Fynbo for allowing us access to their awarded time at the Nordic Optical Telescope, thus making this work possible. We are also grateful to V. J. S. Béjar, E. Guinan, S. Seager, B. Portmann, A. Garcia-Muñoz and Y. Pavlenko for discussions. Support for this project was provided by the Spanish Ministry of Science via the Ramon y Cajal fellowship (E.P.) and project AYA2007-67458. This work was based on observations made with the WHT (operated by the Isaac Newton Group) and the NOT (operated by Denmark, Finland, Iceland, Norway and Sweden), both on the island of La Palma in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.P. (epalle@iac.es).

Existence of collisional trajectories of Mercury, Mars and Venus with the Earth

J. Laskar¹ & M. Gastineau¹

It has been established that, owing to the proximity of a resonance with Jupiter, Mercury's eccentricity can be pumped to values large enough to allow collision with Venus within 5 Gyr (refs 1–3). This conclusion, however, was established either with averaged equations^{1,2} that are not appropriate near the collisions or with non-relativistic models in which the resonance effect is greatly enhanced by a decrease of the perihelion velocity of Mercury^{2,3}. In these previous studies, the Earth's orbit was essentially unaffected. Here we report numerical simulations of the evolution of the Solar System over 5 Gyr, including contributions from the Moon and general relativity. In a set of 2,501 orbits with initial conditions that are in agreement with our present knowledge of the parameters of the Solar System, we found, as in previous studies², that one per cent of the solutions lead to a large increase in Mercury's eccentricity—an increase large enough to allow collisions with Venus or the Sun. More surprisingly, in one of these high-eccentricity solutions, a subsequent decrease in Mercury's eccentricity induces a transfer of angular momentum from the giant planets that destabilizes all the terrestrial planets ~ 3.34 Gyr from now, with possible collisions of Mercury, Mars or Venus with the Earth.

Owing to chaotic behaviour of the Solar System^{4–7}, the distance between two initially close orbital solutions increases by a factor of ten every ten million years^{4,7,8}. It is thus hopeless to search for a precise solution for the motion of the Solar System over 5 Gyr, that is, over a time comparable to its age or life expectancy (before the Sun becomes a red giant). The most precise long-term solutions for the orbital motion of the Solar System are not valid over more than a few tens of millions of years^{9,10}. A numerical integration of the Solar System's motion over 5 Gyr can thus only be considered as a random sample of its possible evolution. Statistical studies are then required to search for possible changes in the planetary orbits that lead to collisions or disruption of the system.

The first study of the planetary orbits over several billion years was obtained after averaging the equations of motion over the fast orbital motion of the planets, which allowed a decrease in integration time of three orders of magnitude¹. This study revealed the possibility of Mercury's eccentricity reaching very high values, allowing collisions with Venus in less than 5 Gyr. However, this method has the drawback that the averaged equations are no longer justified in the vicinity of a collision. Despite the increase in computational power, recent integrations on gigayear timescales still neglected the relativistic contribution^{2,3,11}. The study of the full Solar System using general relativity and non-averaged equations on gigayear timescale thus needed to be done.

The model for the integration of the planetary orbits is derived from the La2004 model⁹ that was integrated over 250 Myr for the study of the palaeoclimates of the Earth and Mars^{9,12}. It comprises the eight major planets and Pluto and includes relativistic¹³ and averaged

lunar contributions¹⁴ (Supplementary Information). We used the SABA4 symplectic integrator¹⁵, which is adapted to perturbed Hamiltonian systems. The step size is 2.5×10^{-2} years, unless the eccentricity of the planets increases beyond about 0.4, in which case the step size is reduced to preserve numerical accuracy.

In a first experiment performed without the Moon or relativistic contributions, we integrated the equations over 5 Gyr for 201 orbits with the same initial conditions, except for an offset of 3.8 km ($k \in [-100, 100]$) in the semi-major axis of Mercury (its actual uncertainty is of a few metres (refs 16, 17)). In 60% of the solutions, we observed large increases in Mercury's eccentricity, beyond 0.9 (Fig. 1a). The statistics (Supplementary Table 1a) are comparable to the results of the non-relativistic secular equations². Among the 121 solutions of large eccentricity ($e > 0.7$), 34 ended in collision with

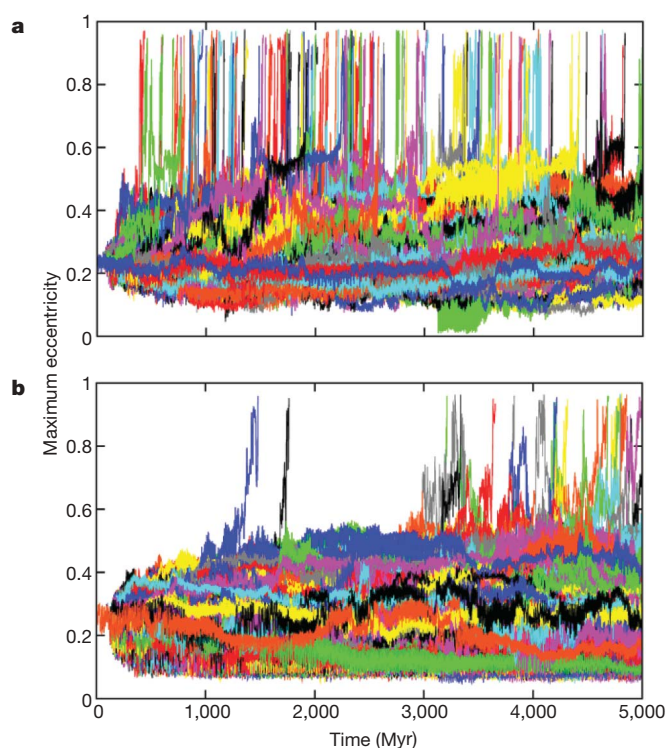


Figure 1 | Mercury's eccentricity over 5 Gyr. Evolution of the maximum eccentricity of Mercury (computed over 1-Myr intervals) over 5 Gyr. **a**, Pure Newtonian model without the contribution of the Moon, for 201 solutions with initial conditions that differ by only 3.8 cm in the semi-major axis of Mercury. **b**, Full Solar System model with relativistic and lunar contributions, for 2,501 solutions with initial conditions that differ by only 0.38 mm in the semi-major axis of Mercury.

¹Astronomie et Systèmes Dynamiques, IMCCE-CNRS UMR8028, Observatoire de Paris, UPMC, 77 Avenue Denfert-Rochereau, 75014 Paris, France.

the Sun, 86 in collision with Venus and a single one reached the 5-Gyr limit before collision.

According to the secular analysis², we could expect a much smaller number of collisional orbits in the full relativistic model. To estimate the probabilities of large eccentricity deviations in the relativistic model, we thus had to increase the scale of the numerical experiment. Using the JADE supercomputer at the French National Computing Centre CINES, we integrated 2,501 orbital solutions, S_0 , of the complete model over 5 Gyr, with the initial semi-major axis of Mercury differing by 0.38 k mm ($k \in [-1,250, 1,250]$) from that in the nominal solution, S_0 , which was adjusted to the planetary ephemeris INPOP06¹⁷. The results (Fig. 1b and Supplementary Table 1b) are comparable to those of the relativistic secular equations², with Mercury having a high eccentricity in about 1% of solutions.

Among these 2,501 solutions that are compatible with our best knowledge of the Solar System, in 20 the eccentricity of Mercury increased beyond 0.9. At the time of writing, 14 of these have not yet reached 5 Gyr and may still be running for a few months, as their step size is greatly reduced. Solution S_{-947} reached 5 Gyr without collision, although a close encounter (closest approach, 6,500 km) occurred between Mercury and Venus at 4.9 Gyr. In S_{-915} , S_{-210} and S_{33} , Mercury collided with the Sun at 4.218, 4.814 and 4.314 Gyr, respectively, whereas in S_{-812} , Mercury collided with Venus at 1.763 Gyr. The most notable solution is S_{-468} , in which a close encounter of Mars with the Earth with a closest approach of only 794 km occurs at 3.3443 Gyr (Fig. 2). Such a close approach would be disastrous for life on the Earth, with a possible tidal disruption of Mars and subsequent multiple impacts on the Earth^{18,19}, but we also searched for more direct collisions. We integrated 201 different versions, $S_{-468}^{(k)}$ ($k \in [-100, 100]$), of S_{-468} , starting at 3.344298 Gyr,

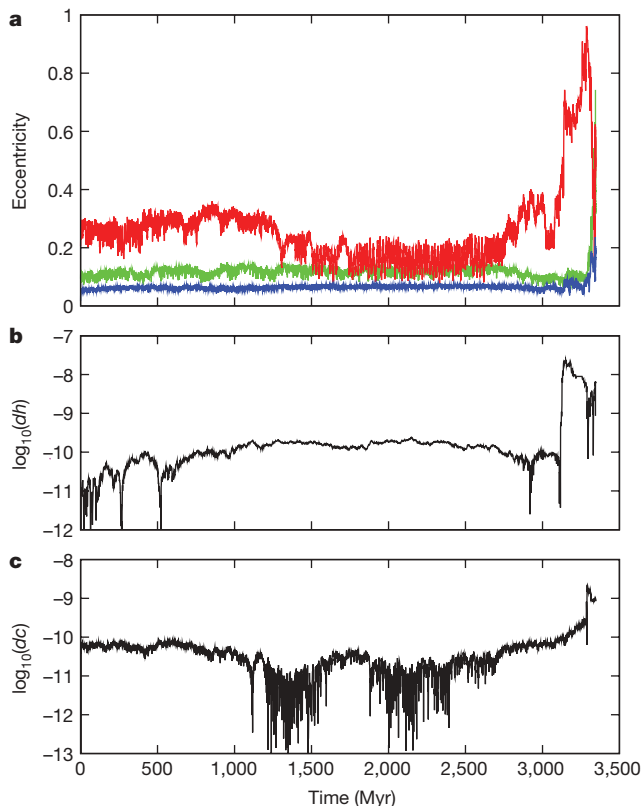


Figure 2 | Example of collisional trajectory for Mars and the Earth. **a**, Evolution of the maximum eccentricity of Mercury (red), Mars (green) and the Earth (blue), recorded over 1-Myr intervals for the solution $S_{-468}^{(-15)}$. **b**, **c**, Relative variation, dh , of the total energy of the system (**b**) and total angular momentum, dc (**c**), on a logarithmic scale. During the whole integration, the relative variations of the total energy and angular momentum of the system remain below 2.3×10^{-8} and 2.2×10^{-9} , respectively.

each with an offset of 0.15 k mm in the semi-major axis of Mars. Within 100 Myr, five cases lead to the ejection of Mars from the Solar System (semi-major axis >100 AU) and the remaining 196 solutions end in collision, with the following distribution: Sun–Mercury, 33; Sun–Mars, 48; Mercury–Venus, 43; Mercury–Earth, 1; Mercury–Mars, 1; Venus–Earth, 18; Venus–Mars, 23; Earth–Mars, 29.

The most surprising collision is the one of Venus with the Earth, which occurs in $S_{-468}^{(-1)}$ in a five-stage process (Figs 2 and 3). The first step is the increase in the eccentricity of Mercury, obtained through perihelion resonance with Jupiter^{2,3} at 3.137 Gyr. This step is essential, as it allows a transfer of non-circular angular momentum from the outer planets to the terrestrial planets²⁰. The eccentricity increase of Venus, the Earth and Mars, is then obtained through secular resonances among the inner planets while the eccentricity of Mercury decreases between 3.305 and 3.325 Gyr. Once Mars and the Earth acquire large eccentricities, close encounters occur and collisions become possible, as in $S_{-468}^{(-15)}$ (Fig. 3c). In $S_{-468}^{(-1)}$, the collision with Mars does not occur, but several close encounters (Fig. 3c) lead to the

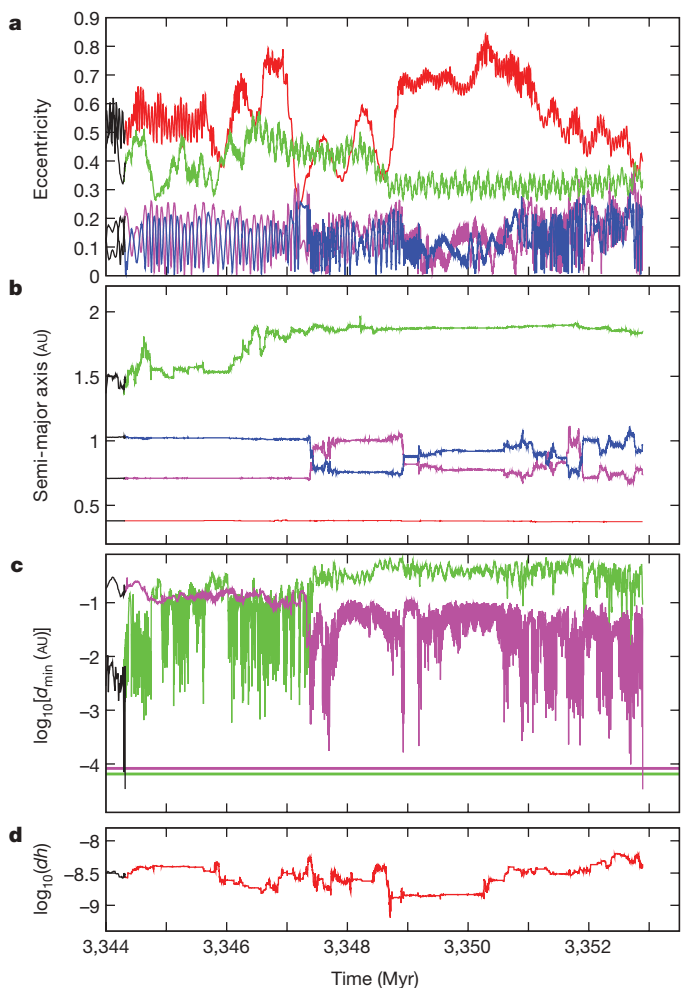


Figure 3 | Collisional trajectories for Mars and Venus with the Earth. The two solutions, $S_{-468}^{(-15)}$ (black) and $S_{-468}^{(-1)}$ (colours), are identical until 3.344298 Gyr. At this time, an offset of 0.15 k mm is applied to the semi-major axis of Mars in the solution $S_{-468}^{(k)}$. **a**, **b**, Eccentricity (**a**) and semi-major axis (**b**) plotted versus time for Mercury (red), Venus (pink), the Earth (blue) and Mars (green). **c**, Minimum Earth–Mars (green) and Earth–Venus (pink) distances in astronomical units, recorded over each 1,000-yr time interval. The horizontal lines are the Mars–Earth (6.5322×10^{-5} AU, green) and Venus–Earth (8.30827×10^{-5} AU, pink) distances of collision, d_{\min} , corresponding to the sum of the planets' radii. For $S_{-468}^{(-15)}$, the Earth–Mars distance reached at 3.344317 Gyr is 3.7752×10^{-5} AU, and for $S_{-468}^{(-1)}$, the Earth–Venus distance reached at 3.352891 Gyr is 3.3791×10^{-5} AU. **d**, Relative energy variation.

diffusion of Mars's semi-major axis (Fig. 3b) until secular resonances produce a decrease in the eccentricity of Mercury together with an additional increase in the eccentricity of Venus and the Earth at about 3.347.3 Gyr (Fig. 3c). At this point, close encounters between Venus and the Earth occur, with several exchanges of the planets' orbits (Fig. 3b) before a final collision at 3.352891 Gyr (Fig. 3c).

The essential trigger for the collisional trajectories of Mars and Venus with the Earth is the great increase in Mercury's eccentricity. If this increase leads rapidly to a collision with the Sun or with Venus, as in S_{33} and S_{-812} , the remaining part of the Solar System is not much affected; these two solutions, after merging of the colliding bodies, are actually very stable when extended to 5 Gyr. We expect that this will be the case for most of the solutions in which Mercury has high eccentricity. However, in some less frequent events, such as S_{-468} , the eccentricity increase of Mercury leads to a total destabilization of the inner Solar System. In the simulations described here, we restricted the initial conditions to a very small neighbourhood, assuming that the chaotic behaviour of the system will randomize the initial conditions. The fact that the probability distribution of Mercury's eccentricity (Supplementary Table 1) is very similar to the results obtained with the averaged equations and a much more widely distributed set of initial conditions² leads us to propose that the probability of a large increase in the eccentricity of Mercury is about 1%. It remains difficult to evaluate the probability of a collision involving the Earth within 5 Gyr. Indeed, although we studied variations around solution S_{-468} that lead to collision with the Earth in about 25% of the cases, the S_{-468} orbit is essentially the single event of its kind in our full sample of 2,501 solutions.

Received 17 February; accepted 22 April 2009.

1. Laskar, J. Large scale chaos in the Solar System. *Astron. Astrophys.* **287**, L9–L12 (1994).
2. Laskar, J. Chaotic diffusion in the Solar System. *Icarus* **185**, 312–330 (2008).
3. Batygin, K. & Laughlin, G. On the dynamical stability of the Solar System. *Astrophys. J.* **683**, 1207–1216 (2008).
4. Laskar, J. A numerical experiment on the chaotic behaviour of the Solar System. *Nature* **338**, 237–238 (1989).
5. Laskar, J. The chaotic motion of the Solar System. A numerical estimate of the size of the chaotic zones. *Icarus* **88**, 266–291 (1990).
6. Laskar, J., Quinn, T. & Tremaine, S. Confirmation of resonant structure in the Solar System. *Icarus* **95**, 148–152 (1992).
7. Sussman, G. J. & Wisdom, J. Chaotic evolution of the Solar System. *Science* **257**, 56–62 (1992).
8. Laskar, J. The limits of Earth orbital calculations for geological time scale use. *Phil. Trans. R. Soc. Lond. A* **357**, 1735–1759 (1999).
9. Laskar, J. et al. A long term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* **428**, 261–285 (2004).
10. Varadi, F., Runnegar, B. & Ghil, M. Successive refinements in long-term integrations of planetary orbits. *Astrophys. J.* **592**, 620–630 (2003).
11. Ito, T. & Tanikawa, K. Long-term integrations and stability of planetary orbits in our Solar System. *Mon. Not. R. Astron. Soc.* **336**, 483–500 (2002).
12. Laskar, J. et al. Long term evolution and chaotic diffusion of the insolation quantities of Mars. *Icarus* **170**, 343–364 (2004).
13. Saha, P. & Tremaine, S. Long-term planetary integration with individual time steps. *Astron. J.* **108**, 1962–1969 (1994).
14. Boué, G. & Laskar, J. Precession of a planet with a satellite. *Icarus* **196**, 1–15 (2008).
15. Laskar, J. & Robutel, P. High order symplectic integrators for perturbed Hamiltonian systems. *Celest. Mech. Dynam. Astron.* **80**, 39–62 (2001).
16. Standish, M. An approximation to the errors in the planetary ephemerides of the Astronomical Almanac. *Astron. Astrophys.* **417**, 1165–1171 (2004).
17. Fienga, A., Manche, H., Laskar, J. & Gastineau, M. INPOP06: a new numerical planetary ephemeris. *Astron. Astrophys.* **477**, 315–327 (2008).
18. Holsapple, K. A. & Michel, P. Tidal disruptions: a continuum theory for solid bodies. *Icarus* **183**, 331–348 (2006).
19. Asphaug, E., Agnor, C. B. & Williams, Q. Hit-and-run planetary collisions. *Nature* **439**, 155–160 (2006).
20. Laskar, J. Large scale chaos and the spacing of the inner planets. *Astron. Astrophys.* **317**, L75–L78 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work benefited from support from the Planetology Programme of the French National Research Centre, from Paris Observatory and from National Research Agency grant ASTS-CM. The authors thank the computing centres of Paris Observatory, the Institut de Physique du Globe Paris, the Institute of Development and Resources in Scientific Computing, the Linear Accelerator Laboratory Grid and especially the French National Computing Centre CINES, for providing the necessary computational resources for this work.

Author Contributions J.L. designed the study, performed the simulations and their analysis, and wrote the paper. M.G. wrote the computer code.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.L. (laskar@imcce.fr).

LETTERS

Direct observation of a widely tunable bandgap in bilayer graphene

Yuanbo Zhang^{1*}, Tsung-Ta Tang^{1*†}, Caglar Girit¹, Zhao Hao^{2,4}, Michael C. Martin², Alex Zettl^{1,3}, Michael F. Crommie^{1,3}, Y. Ron Shen^{1,3} & Feng Wang^{1,3}

The electronic bandgap is an intrinsic property of semiconductors and insulators that largely determines their transport and optical properties. As such, it has a central role in modern device physics and technology and governs the operation of semiconductor devices such as p–n junctions, transistors, photodiodes and lasers¹. A tunable bandgap would be highly desirable because it would allow great flexibility in design and optimization of such devices, in particular if it could be tuned by applying a variable external electric field. However, in conventional materials, the bandgap is fixed by their crystalline structure, preventing such bandgap control. Here we demonstrate the realization of a widely tunable electronic bandgap in electrically gated bilayer graphene. Using a dual-gate bilayer graphene field-effect transistor (FET)² and infrared microspectroscopy^{3–5}, we demonstrate a gate-controlled, continuously tunable bandgap of up to 250 meV. Our technique avoids uncontrolled chemical doping^{6–8} and provides direct evidence of a widely tunable bandgap—spanning a spectral range from zero to mid-infrared—that has eluded previous attempts^{2,9}. Combined with the remarkable electrical transport properties of such systems, this electrostatic bandgap control suggests novel nanoelectronic and nanophotonic device applications based on graphene.

Graphene's unique electronic band structure has led to fascinating phenomena, exemplified by massless Dirac fermion physics^{10–12} and an anomalous quantum Hall effect^{13–16}. With one more graphene layer added, bilayer graphene has an entirely different (and equally interesting) band structure. Most notably, the inversion symmetric AB-stacked bilayer graphene is a zero-bandgap semiconductor in its pristine form. But a non-zero bandgap can be induced by breaking the inversion symmetric of the two layers. Indeed, a bandgap has been observed in a one-side chemically doped epitaxial graphene bilayer^{6,8}.

Of particular importance, however, is the potential of a continuously tunable bandgap through an electrical field applied perpendicularly to the sample^{17–20}. Such control has proven elusive. Electrical transport measurements on gated bilayer exhibit insulating behaviour only at temperatures below 1 kelvin², suggesting a bandgap value much lower than theoretical predictions^{17,18}. Optical studies of bilayers have so far been limited to samples with a single electrical gate^{4,5,9}, in which carrier doping effects dominate and obscure the signatures of a gate-induced bandgap. Such lack of experimental evidence has cast doubt on the possibility of achieving gate-controlled bandgaps in graphene bilayers⁹.

Here, we use novel dual-gate graphene FETs to demonstrate unambiguously a widely field-tunable bandgap in bilayer graphene with infrared absorption spectroscopy. By using both top and bottom gates in the graphene FET device we are able to control independently the

two key semiconductor parameters: electronic bandgap and carrier doping concentration.

The electronic structure near the Fermi level of an AB-stacked graphene bilayer features two nearly parallel conduction bands above two nearly parallel valence bands (Fig. 1d)²¹. In the absence of gating, the lowest conduction band and highest valence band touch each other with a zero bandgap. Upon electrical gating, the top and bottom electrical displacement fields D_t and D_b (Fig. 1c) produce two effects (Fig. 1d): The difference of the two, $\delta D = D_b - D_t$, leads to a net carrier doping, that is, a shift of the Fermi energy (E_F). The average of the two, $\bar{D} = (D_b + D_t)/2$, breaks the inversion symmetry of the bilayer and generates a non-zero bandgap $D^{7,17,18}$. By setting δD to zero and varying \bar{D} , we can tune the bandgap while keeping the bilayer charge neutral. Sets of D_b and D_t leading to $\delta D = 0$ define the bilayer 'charge neutral points' (CNPs). By varying δD above or below zero, we can inject electrons or holes into the bilayer and shift the Fermi level without changing the bandgap. In our experiment the drain electrode is grounded and the displacement fields D_t and D_b are tuned independently by top and bottom gate voltages (V_t and V_b) through the relations $D_b = +\epsilon_b(V_b - V_b^0)/d_b$ and $D_t = -\epsilon_t(V_t - V_t^0)/d_t$. Here ϵ and d are the dielectric constant and thickness of the dielectric layer and V^0 is the effective offset voltage due to initial environment-induced carrier doping.

The relationship between D and V for the top or bottom layers can be determined through electrical transport measurement². Figure 1e shows the measured resistance along the graphene plane as a function of V_t with V_b fixed at different values, and CNPs can be identified by the peaks in the resistance curves, because charge neutrality results in a maximum resistance. The deduced CNPs, in terms of (V_t, V_b) , are plotted in Fig. 1f. V_t and V_b are linearly related with a slope of about 0.15, consistent with the expected value of $-(\epsilon_b d_t)/(\epsilon_t d_b)$, where $d_b = 285$ nm, $\epsilon_b = 3.9$ for thermal SiO₂, and $d_t = 80$ nm, $\epsilon_t = 7.5$ for amorphous Al₂O₃. The peak resistance differs at different CNPs (Fig. 1e) because the field-induced bandgap itself differs. Lower peak resistance comes from a smaller bandgap. Thus, the lowest peak resistance allows us roughly to identify the zero-bandgap CNP ($D_b = D_t = 0$) and determine the offset top and bottom gate voltages from environment doping to be $V_t^0 \approx -5$ V and $V_b^0 \approx 10$ V. With the values of ϵ/d and gate voltage offsets, the displacement electric field can be determined within an uncertainty of about 10%. We note that although CNP resistance data shows an increase with the field-induced bandgap, the increase is much smaller than expected for a large energy gap opening. This is attributed to extrinsic conduction through defects and carrier doping from charge impurities in our samples.

To determine the true bilayer bandgap reliably, we used infrared microspectroscopy^{3,4}. Such an optical determination of the electronic

¹Department of Physics, University of California at Berkeley, ²Advanced Light Source Division, Lawrence Berkeley National Laboratory, ³Materials Science Division, Lawrence Berkeley National Laboratory, ⁴Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. †Present address: Department of Photonics and Institute of Electro-optical Engineering, National Chiao Tung University, Hsinchu, Taiwan 30010.

*These authors contributed equally to this work.

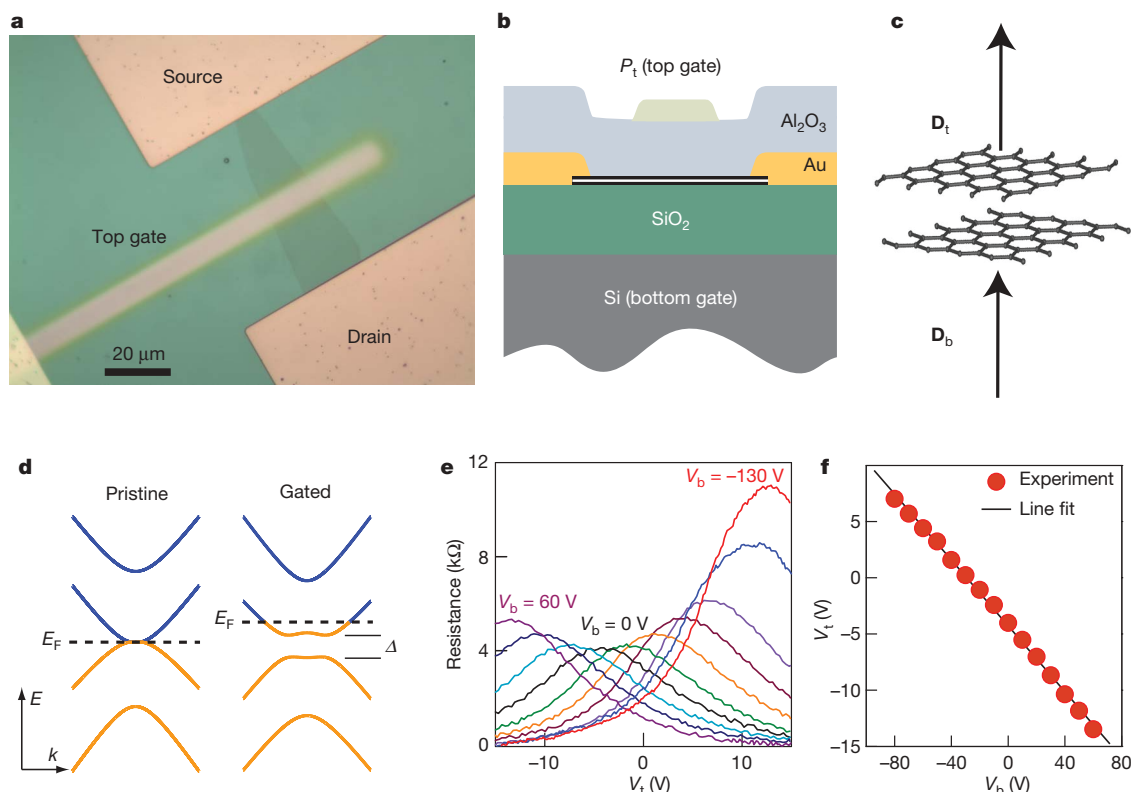


Figure 1 | Dual-gated bilayer graphene. **a**, Optical microscopy image of the bilayer device (top view). **b**, Illustration of a cross-sectional side view of the gated device. **c**, Sketch showing how gating of the bilayer induces top (D_t) and bottom (D_b) electrical displacement fields. **d**, Left, the electronic structure of a pristine bilayer has zero bandgap. (k denotes the wavevector.) Right, upon gating, the displacement fields induces a non-zero bandgap Δ .

bandgap is generally less affected by defects or doping than electrical transport measurements². Figure 2b shows the gate-induced bilayer absorption spectra at CNPs ($\delta D = 0$) with $\bar{D} = 1.0 \text{ V nm}^{-1}$, 1.4 V nm^{-1} , 1.9 V nm^{-1} and 3.0 V nm^{-1} . The absorption spectrum of the sample at the zero-bandgap CNP ($\bar{D} = 0$) has been subtracted as a background reference to eliminate contributions to the absorption

and a shift of the Fermi energy E_F . **e**, Graphene electrical resistance as a function of top gate voltage V_t at different fixed bottom gate voltages V_b . The traces are taken with 20 V steps in V_b from 60 V to -100 V and at $V_b = -130 \text{ V}$. The resistance peak in each curve corresponds to the CNP ($\delta D = 0$) for a given V_b . **f**, The linear relation between top and bottom gate voltages that results in bilayer CNPs.

from the substrate and gate materials. Two distinct features are present in the spectra, a gate-dependent peak below 300 meV and a dip centred around 400 meV. These arise from different optical transitions between the bilayer electronic bands, as illustrated in Fig. 2a. Transition I is the tunable bandgap transition that accounts for the gate-induced spectral response at energies lower than 300 meV. Transitions II, III, IV and V

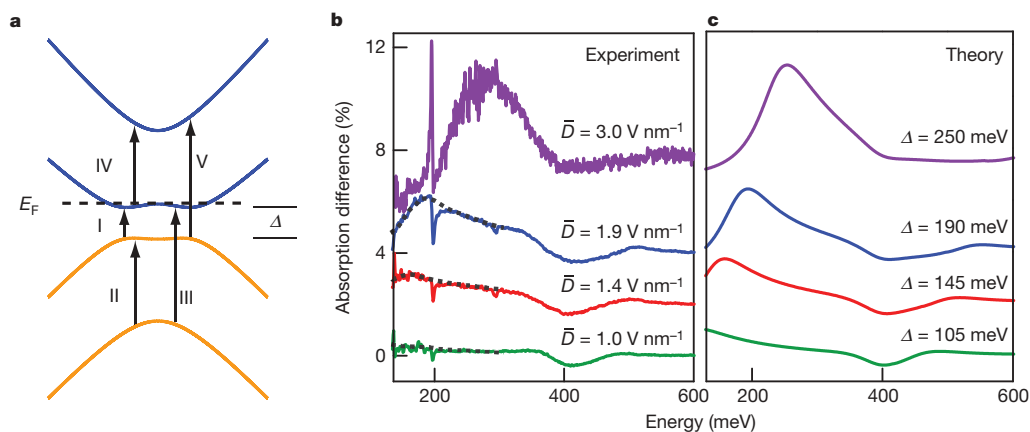


Figure 2 | Bilayer energy gap opening at strong electrical gating. **a**, Allowed optical transitions between different sub-bands of a graphene bilayer. Curves are offset from zero for clarity. **b**, Gate-induced absorption spectra at CNP for different applied displacement fields \bar{D} (with the spectrum for zero-bandgap CNP subtracted as reference). For clarity the upper traces were displaced by 2%, 4% and 8%, respectively. Absorption peaks due to transition I at gate-induced bandgaps are apparent (dashed black lines are guides to the eye). At the same time, a reduction of absorption below the bandgap is expected. This reduction is clearly observed in the trace with the

largest bandgap ($\Delta = 250 \text{ meV}$) in our experimental spectral range. The sharp asymmetric resonance observed near 200 meV is due to Fano resonance of the zone-centre G-mode phonon with the continuum electronic transitions. The broad feature around 400 meV is due to electronic transitions II, III, IV and V. **c**, Theoretical prediction of the gate-induced absorption spectra based on a tight-binding model where the bandgap value is taken as an adjustable parameter. The fit provides an accurate determination of the gate-tunable bandgap at strong electrical gating.

occur at and above the energy of parallel band separation ($\gamma \equiv 400$ meV) and contribute to the spectral feature near 400 meV.

The absorption peak below 300 meV in Fig. 2b shows pronounced gate tunability: it gets stronger and shifts to higher energy with increasing \bar{D} . This arises because as the bandgap increases, so does the density of states at the band edge. The peak position, corresponding to the bandgap, increases from 150 meV at $\bar{D} = 1.4$ V nm⁻¹ to 250 meV at $\bar{D} = 3$ V nm⁻¹. This shows directly that the bandgap can be continuously tuned up to at least 250 meV by electrical gating. The bandgap transitions are remarkably strong: optical absorption can reach 5% in two atom layers, corresponding to an oscillator strength that is among the highest of all known materials. On the basis of the sum rule, a reduction of absorption below the bandgap should accompany the prominent band-edge absorption peak. This absorption reduction is clearly observed in the trace with the largest bandgap ($\Delta = 250$ meV) in our experimental spectral range. We also notice in Fig. 2b a very sharp spectral feature at 1,585 cm⁻¹ (about 200 meV). This narrow resonance can be attributed to the zone-centre G-mode phonon in graphene²². The asymmetric lineshape originates from Fano interference between the discrete phonon and continuous electronic (bandgap) transitions.

When the displacement field \bar{D} is weak (< 1.2 V nm⁻¹), the gate-induced bandgap becomes too small to be measured directly. However, it can still be extracted from spectral changes around 400 meV induced by electron doping through gating. This is achieved by measuring the difference in bilayer absorption for $\delta D = 0$ (CNP) and $\delta D = 0.15$ V nm⁻¹ (electron-doped) at different fixed \bar{D} values (Fig. 3a). We first examine the optical transitions in Fig. 2a, to understand the bilayer absorption difference due to electron doping. With electrons occupying the conduction band states, transition IV becomes stronger from extra filled initial states and transition III becomes weaker because of fewer available empty final states. However, transition IV is more prominent and gives rise to the observed peaks in the absorption difference spectra because all such transitions have similar energy owing to the nearly parallel conduction

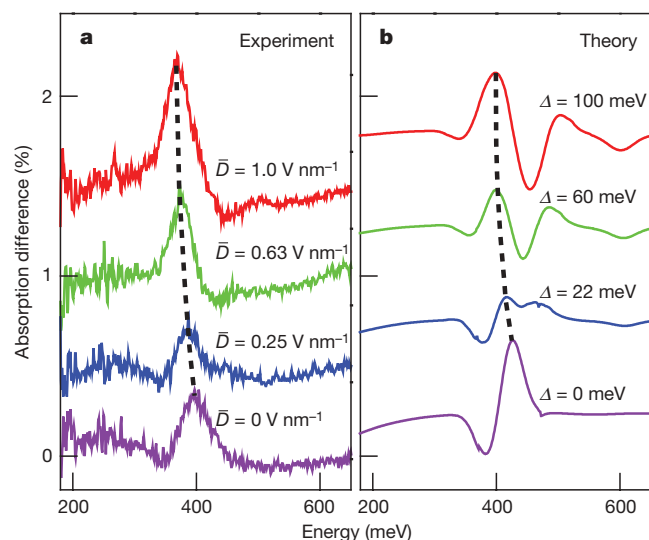


Figure 3 | Bilayer energy gap opening at weak electrical gating. **a**, Absorption difference between electron-doped ($\delta D = 0.15$ V nm⁻¹) and charge-neutral bilayers ($\delta D = 0$) at different average displacement fields \bar{D} . The curves are displaced by multiples of 0.5% for clarity. The absorption peak is mainly due to increased absorption between nearly parallel conduction bands from extra filled initial states (transition IV in Fig. 2a). This absorption peak shifts to lower energy due to the opening of the bilayer bandgap with increasing \bar{D} . **b**, Calculated absorption difference spectra based on a tight-binding model using the gate-induced bandgap as an adjustable parameter. Good agreement between theory and experiment on the absorption peak redshift (black dashed lines in **a** and **b**) yields the gate-induced bilayer bandgap at weak gating.

bands. When the bandgap increases with increasing \bar{D} , the lower conduction band moves up, but the upper conduction band hardly changes, making the separation between the two bands smaller. This will lead to a redshift of transition IV. Therefore, the shift of the peak in the difference spectrum can yield the bilayer bandgap when compared to theory. When the gate-induced bandgap is small, this shift equals roughly half of the bandgap energy. At higher \bar{D} values, deviation from the near-parallel band picture becomes significant and a broadening of the absorption peak takes place (Supplementary Fig. S1).

We obtained quantitative understanding of the gate-induced bandgap and its associated optical properties through comparison of our data to theoretical predictions. We modelled the bilayer absorption using the self-consistent tight-binding model following ref. 23, except that the bandgap was treated as a fitting parameter here. We have included a room-temperature thermal broadening of 25 meV and an extra inhomogeneous broadening of 60 meV to account for sample inhomogeneity. We note that this large inhomogeneous broadening is comparable to that estimated from transport studies²⁴ and it accounts for the difficulty in electrical determination of the bilayer graphene bandgap. Figure 2c shows our calculated gate-induced absorption spectra and bandgaps of bilayer graphene extracted by matching the absorption peak between 130–300 meV in the ‘large bandgap’ regime ($\Delta > 120$ meV). Agreement with the experimental spectra (Fig. 2b) is excellent, except for the phonon contribution at ~ 200 meV, which is not included in our model. For the ‘small bandgap’ regime ($\Delta < 120$ meV), we are able to determine the bilayer bandgap by comparing our model calculations to the measured absorption difference spectra shown in Fig. 3a. Our calculations (Fig. 3b) provide a good qualitative fit to the absorption peak that arises from electron transition IV: this absorption peak shifts to lower energy as the bandgap becomes larger, reproducing the observed behaviour at increasing displacement field \bar{D} in Fig. 3a. By matching the experimental and theoretical values of this absorption peak shift, we can extract the bilayer bandgap at different \bar{D} in the ‘small bandgap’ regime.

Figure 4 shows a plot of the experimentally derived gate-tunable bilayer bandgap over the entire range ($0 < \Delta < 250$ meV) as a function of applied displacement field \bar{D} (data points). Our experimental bandgap results are compared to predictions based on self-consistent tight-binding calculations (black trace)²³, *ab initio* density functional (red trace)¹⁸, and unscreened tight-binding calculations (dashed blue line)⁷. Clearly the inclusion of graphene self-screening is crucial in achieving good agreement with the experimental data, as in the self-consistent tight-binding and *ab initio* calculations. The *ab initio*

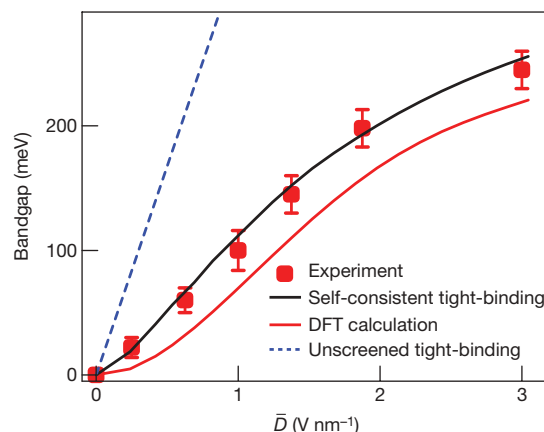


Figure 4 | Electric-field dependence of tunable energy bandgap in graphene bilayer. Experimental data (red squares) are compared to theoretical predictions based on self-consistent tight-binding (black trace), *ab initio* density functional (red trace), and unscreened tight-binding calculations (blue dashed trace). The error bar is estimated from the uncertainty in determining the absorption peaks in the spectra.

calculation predicts a slightly smaller bandgap than does the tight-binding model. This is partly owing to the different values used for onsite interlayer coupling γ_1 , which is 0.4 eV for the tight binding and 0.34 eV for the *ab initio* calculations. Similar underestimation of bandgaps by *ab initio* local density functional calculations is common for semiconductors²⁵.

Our study shows a confluence of interesting electronic and optical properties in graphene bilayer FETs, which provide appealing opportunities for new scientific exploration and technological innovation. The achieved gate-tunable bandgap (250 meV), an order of magnitude higher than the room-temperature thermal energy (25 meV), emphasizes the intrinsic potential of bilayer graphene for nanoelectronics. With the tunable bandgap reaching the infrared range, and with the unusually strong oscillator strength for the bandgap transitions, bilayer graphene may enable novel nanophotonic devices for infrared light generation, amplification and detection.

METHODS SUMMARY

Graphene bilayer flakes were exfoliated from graphite and deposited onto Si/SiO₂ wafers as described in ref. 26. Bilayers were identified by optical contrast in a microscope and subsequently confirmed via Raman spectroscopy²². Source and drain electrodes (Au, thickness 30 nm) for transport measurement were deposited directly onto the graphene bilayer through a stencil mask under vacuum. The doped Si substrate under a 285-nm-thick SiO₂ layer was used as the bottom gate. The top gate was formed by sequential deposition of an 80-nm-thick Al₂O₃ film and a sputtered strip of 20-nm-thick Pt film. The Pt electrode was electrically conductive and optically semi-transparent. Two-terminal electrical measurements were used for transport characterization. We extracted a carrier mobility of $\sim 1,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ from the electrical transport measurements. Infrared transmission spectra of the dual-gated bilayer were obtained using the synchrotron infrared beamline at the Advanced Light Source at Berkeley and a microFourier transform infrared spectrometer. All measurements were performed at room temperature (293 K).

Received 26 February; accepted 30 April 2009.

1. Sze, S. M. & Ng, K. K. *Physics of Semiconductor Devices* (Wiley-Interscience, 2006).
2. Oostinga, J. B., Heersche, H. B., Liu, X. L., Morpurgo, A. F. & Vandersypen, L. M. K. Gate-induced insulating state in bilayer graphene devices. *Nature Mater.* **7**, 151–157 (2008).
3. Li, Z. Q. *et al.* Dirac charge dynamics in graphene by infrared spectroscopy. *Nature Phys.* **4**, 532–535 (2008).
4. Wang, F. *et al.* Gate-variable optical transitions in graphene. *Science* **320**, 206–209 (2008).
5. Li, Z. Q. *et al.* Band structure asymmetry of bilayer graphene revealed by infrared spectroscopy. *Phys. Rev. Lett.* **102**, 037403 (2009).
6. Ohta, T., Bostwick, A., Seyller, T., Horn, K. & Rotenberg, E. Controlling the electronic structure of bilayer graphene. *Science* **313**, 951–954 (2006).
7. Castro, E. V. *et al.* Biased bilayer graphene: semiconductor with a gap tunable by the electric field effect. *Phys. Rev. Lett.* **99**, 216802 (2007).

8. Zhou, S. Y. *et al.* Substrate-induced bandgap opening in epitaxial graphene. *Nature Mater.* **6**, 770–775 (2007).
9. Kuzmenko, A. B. *et al.* Infrared spectroscopy of electronic bands in bilayer graphene. Preprint at <<http://arxiv.org/abs/0810.2400>> (2008).
10. Geim, A. K. & Novoselov, K. S. The rise of graphene. *Nature Mater.* **6**, 183–191 (2007).
11. Katsnelson, M. I., Novoselov, K. S. & Geim, A. K. Chiral tunnelling and the Klein paradox in graphene. *Nature Phys.* **2**, 620–625 (2006).
12. Huard, B. *et al.* Transport measurements across a tunable potential barrier in graphene. *Phys. Rev. Lett.* **98**, 236803 (2007).
13. Novoselov, K. S. *et al.* Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438**, 197–200 (2005).
14. Zhang, Y. B., Tan, Y. W., Stormer, H. L. & Kim, P. Experimental observation of the quantum Hall effect and Berry's phase in graphene. *Nature* **438**, 201–204 (2005).
15. Novoselov, K. S. *et al.* Unconventional quantum Hall effect and Berry's phase of 2π in bilayer graphene. *Nature Phys.* **2**, 177–180 (2006).
16. McCann, E. & Fal'ko, V. I. Landau-level degeneracy and quantum hall effect in a graphite bilayer. *Phys. Rev. Lett.* **96**, 086805 (2006).
17. McCann, E. Asymmetry gap in the electronic band structure of bilayer graphene. *Phys. Rev. B* **74**, 161403 (2006).
18. Min, H. K., Sahu, B., Banerjee, S. K. & MacDonald, A. H. Ab initio theory of gate induced gaps in graphene bilayers. *Phys. Rev. B* **75**, 155115 (2007).
19. Lu, C. L., Chang, C. P., Huang, Y. C., Chen, R. B. & Lin, M. L. Influence of an electric field on the optical properties of few-layer graphene with AB stacking. *Phys. Rev. B* **73**, 144427 (2006).
20. Guinea, F., Neto, A. H. C. & Peres, N. M. R. Electronic states and Landau levels in graphene stacks. *Phys. Rev. B* **73**, 245426 (2006).
21. Abergel, D. S. L. & Fal'ko, V. I. Optical and magneto-optical far-infrared properties of bilayer graphene. *Phys. Rev. B* **75**, 155430 (2007).
22. Ferrari, A. C. *et al.* Raman spectrum of graphene and graphene layers. *Phys. Rev. Lett.* **97**, 187401 (2006).
23. Zhang, L. M. *et al.* Determination of the electronic structure of bilayer graphene from infrared spectroscopy. *Phys. Rev. B* **78**, 235408 (2008).
24. Adam, S. & Sarma, S. D. Boltzmann transport and residual conductivity in bilayer graphene. *Phys. Rev. B* **77**, 115436 (2007).
25. Hybertsen, M. S. & Louie, S. G. Electron correlation in semiconductors and insulators—band-gaps and quasi-particle energies. *Phys. Rev. B* **34**, 5390–5413 (1986).
26. Novoselov, K. S. *et al.* Two-dimensional atomic crystals. *Proc. Natl Acad. Sci. USA* **102**, 10451–10453 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Office of Basic Energy Sciences, US Department of Energy under contract DE-AC03-76SF0098 (Materials Science Division) and contract DE-AC02-05CH11231 (Advanced Light Source). F.W., Y.Z. and T.-T.T. acknowledge support from a Sloan fellowship, a Miller fellowship and a fellowship from the National Science Council of Taiwan, respectively.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.W. (fengwang76@berkeley.edu).

LETTERS

Total synthesis of eudesmane terpenes by site-selective C–H oxidations

Ke Chen¹ & Phil S. Baran¹

From menthol to cholesterol to Taxol, terpenes are a ubiquitous group of molecules (over 55,000 members isolated so far) that have long provided humans with flavours, fragrances, hormones, medicines and even commercial products such as rubber¹. Although they possess a seemingly endless variety of architectural complexities, the biosynthesis of terpenes often occurs in a unified fashion as a ‘two-phase’ process^{2,3}. In the first phase (the cyclase phase), simple linear hydrocarbon phosphate building blocks are stitched together by means of ‘prenyl coupling’, followed by enzymatically controlled molecular cyclizations and rearrangements. In the second phase (the oxidase phase), oxidation of alkenes and carbon–hydrogen bonds results in a large array of structural diversity. Although organic chemists have made great progress in developing the logic^{3–5} needed for the cyclase phase of terpene synthesis, particularly in the area of polyene cyclizations⁶, much remains to be learned if the oxidase phase is to be mimicked in the laboratory. Here we show how the logic of terpene biosynthesis has inspired the highly efficient and stereocontrolled syntheses of five oxidized members of the eudesmane family of terpenes in a modicum of steps by a series of simple carbocycle-forming reactions followed by multiple site-selective inter- and intramolecular carbon–hydrogen oxidations. This work establishes an intellectual framework in which to conceive the laboratory synthesis of other complex terpenes using a ‘two-phase’ approach.

Synthetic organic chemists have been captivated by terpenes for over 100 years, with the first total syntheses of camphor and terpineol being completed in 1903 and 1904, respectively (ref. 1 and references therein). The origin of terpenes was demystified by the Ruzicka school in 1953 with the formulation of the ‘biogenetic isoprene rule’ (see refs 7 (last chapter, co-authored with A. Eschenmoser and H. Heusser) and 8). Advances in analytical techniques and the development of retrosynthetic analysis⁴ has led to a significant increase in both the number of newly discovered terpenes and the ability to rationally plan and execute their total synthesis^{1,3–5}. However, in the case of complex terpenes such as the clinically used anticancer agent paclitaxel, or Taxol (**2**; Fig. 1), the efficiency of biosynthesis is generally far ahead of the current capabilities of chemical synthesis. For example, tonne quantities of 10-deacetyl baccatin III (**3**) are generated annually (presumably via taxadiene (**1**))⁹ by the needles of the European yew tree as a fully oxidized renewable starting material for the commercial semi-synthesis of **2** (ref. 10), whereas only milligram quantities of totally synthetic **2** are accessible, and that only with significant effort^{1,3}. The lack of general methods, strategies and rules for the functionalization of C–H bonds within complex hydrocarbon systems might account for this difference. Indeed, the total synthesis of multiply oxygenated terpenes using sequential, site-selective C–H functionalizations has not been reported, despite advances in the field of C–H oxidation, specifically for the conversion of a single C–H bond to a C–O bond in diverse settings^{11–17}. Here we

report such total syntheses using an atypical approach to synthesis design that is inspired by terpene biosynthesis and proceeds in two separate phases (Fig. 1).

The eudesmane family of terpenes (Fig. 1), containing over 1,000 members with almost every conceivable oxidation pattern expressed, was selected for this study¹⁸. Despite their low molecular weight, the rigid skeleton of these natural products render them difficult targets for synthesis, especially if large quantities are desired. Dihydrojunenol (**4**), one of the lowest oxidized members, has been prepared only once by a stereorandom semi-synthesis from santonin that proceeded in nine steps (8% overall yield) as a separable mixture of three stereoisomers¹⁹. An elegant enantioselective total synthesis of eudesmanes similar in complexity to **4** proceeded in 15 steps (milligram quantities, 14% overall yield)²⁰. Related family members that have not been previously synthesized are 4-epiajanol (**5**)²¹ dihydroxyeudesmane (**6**)²², pygmal (**7**)²³ and eudesmantetraol (**8**)²⁴; these were thought to arise from **4**. A robust and simple route to **4** (the equivalent of a cyclase phase) would set the stage for site-selective oxidations of two of its five tertiary C–H bonds to access **5–8** (the equivalent of an oxidase phase).

The gram-scale preparation of **4** (Fig. 2) involved a nine-step sequence, five steps of which generated C–C bonds and four of which set key stereocentres. Thus, an enantioselective intermolecular Michael reaction of 3-methyl butyraldehyde and methyl vinyl ketone, catalysed by prolinol **9** (refs 25, 26), followed by treatment with base, led to formation of the natural product cryptone (**10**) in 63% overall yield and 89% enantiomeric excess. The route to this simple terpene is superior in terms of step count and overall yield to a dozen previously reported syntheses (see Supplementary Information for references). Iodination of **10** proceeded in nearly quantitative yield to afford the iodoenone **11**. Introduction of the side chain by Grignard addition and subsequent 1,3-carbonyl transposition furnished intermediate **12** (74% overall yield). The decalin framework of the eudesmanes was forged using an intramolecular Heck cyclization under standard conditions to deliver diene **13** in 95% yield. The final carbon atom was installed by a regio- and stereoselective 1,4-addition and tandem isomerization to afford enone **14** in 56% yield (along with recovered **13**). A two-step strategic reduction sequence set the three consecutive steric centres of **4** in 87% overall yield as a single diastereomer. By this route, gram quantities of enantiopure **4** were prepared (21% overall yield).

With ample quantities of **4** available, a systematic exploration of oxidation chemistry could take place (Fig. 3). Our studies were influenced by those of Chen and White, who were the first to explicitly demonstrate that steric and electronic factors can be used to predict the selectivity of intermolecular tertiary C–H oxidation in complex settings using iron catalysis¹⁷. To begin, the trifluoroethyl carbamate directing group²⁷ was installed in quantitative yield to deliver **15**. This group served the purpose of depleting electron density from the neighbouring carbon atoms (deactivating them against electrophilic

¹Department of Chemistry, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California 92037, USA.

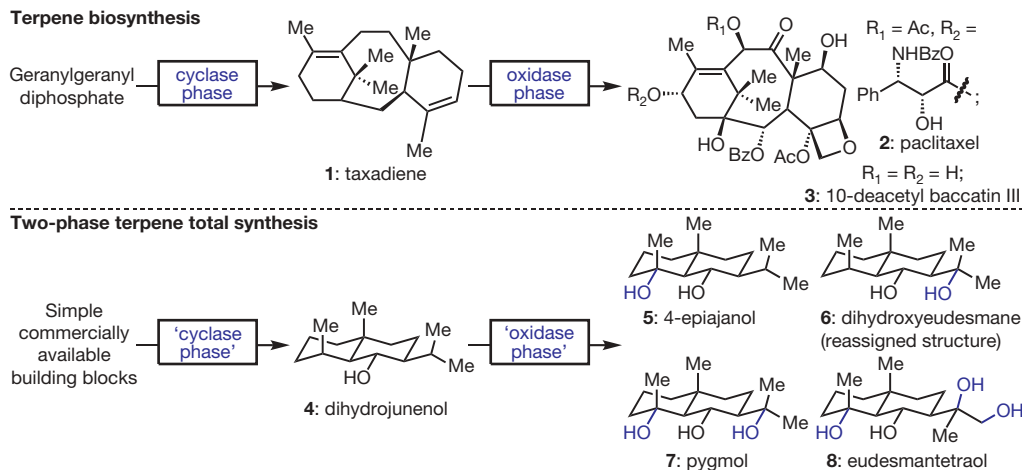


Figure 1 | Outline of the 'two-phase' approach to terpene total synthesis. Me, methyl; Ac, acetyl; Bz, benzoyl.

attack) and would be used to direct other crucial transformations. A C–H bond reactivity analysis using X-ray crystallography and ¹³C nuclear magnetic resonance (NMR) allowed for a qualitative differentiation of the five tertiary C–H bonds in structure **15**, H₁–H₅. Using ¹³C NMR chemical shifts (δ), the relative electronegativity trend for the tertiary carbon atoms was established as follows:

$$\delta_{C3} = 73.6 \text{ p.p.m.} > \delta_{C2} = 55.2 \text{ p.p.m.} \approx \delta_{C4} =$$

$$50.2 \text{ p.p.m.} > \delta_{C1} = 27.5 \text{ p.p.m.} \approx \delta_{C5} = 26.6 \text{ p.p.m.}$$

This implies that H₁ and H₅ are the most likely tertiary C–H bonds to be oxidized with an electrophilic oxidant. Furthermore, X-ray crystallography and modelling studies suggest that H₁ adopts an equatorial orientation and that H₅ is populated by multiple conformers. The studies of Curci¹¹ demonstrated that powerful organic oxidants such as dioxiranes selectively oxidize equatorially oriented C–H bonds in preference to those adopting an axial configuration.

In the event, site-specific intermolecular oxidation of carbamate **15** using TFDO (1.0 equiv.) led to the formation of intermediate **16** in 82% isolated yield on a gram scale (Fig. 3). The origin of this selectivity may stem from strain-release effects in the transition state during oxidation²⁸, a subject to be examined in future investigations. Basic hydrolysis of **16** led to 4-epiajanol (**5**) and its structure was verified by

X-ray crystallography²¹. Consistent with the electronic rules set forth by White and Chen¹⁷, substrate **15** did not incur substantial levels of oxidation at either H₁ or H₅ using Fe-based catalysis. To oxidize H₅ first without disturbing H₁, the trifluoroethyl carbamate directing group was exploited because geometric constraints (as judged from molecular models) prevent this directing group from reaching H₁. Using our previously reported conditions²⁷, site-specific halogenation of **15** to **17** was accomplished followed by cyclization and hydrolysis to afford a compound (**6**) that exhibited spectroscopic measurements identical to those reported for dihydroxyeudesmane (**6'**)²². The originally assigned structure (**6'**) must therefore be reassigned as depicted for **6** (verified by X-ray crystallography).

To access the higher oxidation states of this family (Fig. 4), which contain three or four hydroxyl groups, **16** was oxidized to bromide **18** by using our previously reported conditions to achieve site-specific halogenation. Silver-assisted cyclization to carbonate **19** and hydrolysis afforded natural product pygmal (**7**). This sequence (52% isolated yield in addition to 30% recovered **16**) was carried out in two separate vessels and could be intercepted at the crystalline carbonate **19** to verify the structure by X-ray crystallography. The tetrahydroxylated natural product eudesmantetraol (**8**) was accessed by a formal remote dehydrogenation process. Thus, the tertiary bromide in **18** was eliminated with tetramethylpiperidine to afford an intermediate alkene (**20**). The

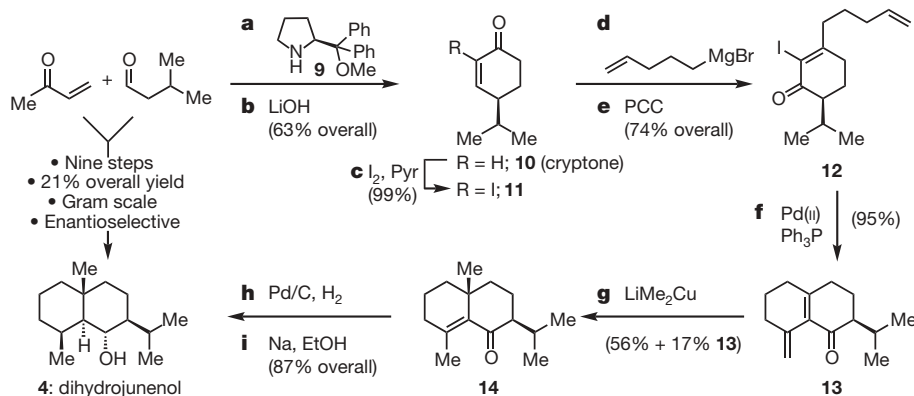


Figure 2 | Simple, enantioselective total synthesis of dihydrojunenol (**4**). Reagents and conditions as follows. **a**, Methyl vinyl ketone (1.5 equiv.), 3-methyl butyraldehyde (1.0 equiv.), prolinol catalyst (0.05 equiv.), ethyl 3,4-dihydroxybenzoate (0.20 equiv.), neat, 4 °C, 36 h, 89%. **b**, LiOH (0.1 equiv.), *i*-PrOH, room temperature (RT, 23 °C), 24 h, 63% over two steps, 89% enantiomeric excess. **c**, I₂ (1.2 equiv.), Pyr/DCM, RT, 12 h, 99%. **d**, (CH₂CHCH₂CH₂CH₂)MgBr (1.5 equiv.), toluene, –78 °C, 30 min; then 0 °C, 30 min. **e**, PCC (1.2 equiv.), 3 Å MS, DCM, RT, 6 h, 74% over two steps. **f**, Pd(OAc)₂ (0.1 equiv.), Ph₃P (0.3 equiv.), Et₃N (1.2 equiv.), Ag₂CO₃

(1.0 equiv.), CH₃CN, 70 °C, 3 h, 95%. **g**, LiMe₂Cu (1.5 equiv.), DCM, 0 °C, 4 h, 56% (17% recovered starting material). **h**, H₂ (1 atm), Pd/C (0.1 equiv.), EtOAc, RT, 30 min. **i**, Na (5 equiv.), EtOH, RT, 30 min, 87% over two steps. Et₃N, triethylamine; DCM, dichloromethane; I₂, iodine; Pyr, pyridine; PCC, pyridinium chlorochromate; MS, molecular sieves; Ph₃P, triphenylphosphine; CH₃CN, acetonitrile; LiMe₂Cu, lithium dimethylcuprate; EtOAc, ethyl acetate. For selected physical data for compounds **11**, **12**, **13**, **14** and **4**, see the Supplementary Information.

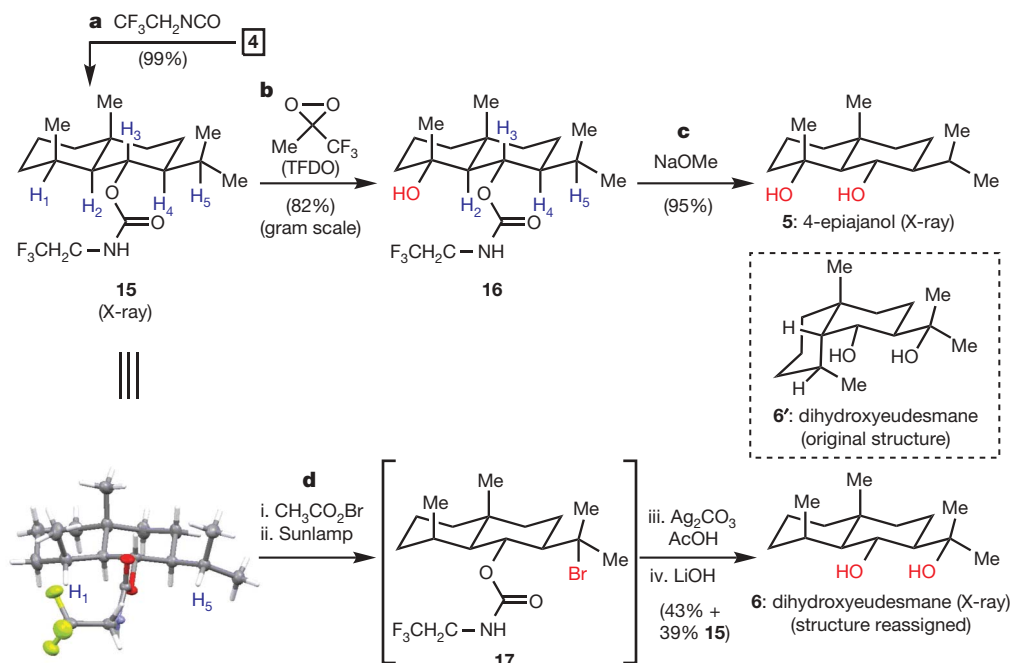


Figure 3 | Total syntheses of 4-epiajanol (**5**) and dihydroxyeudesmane (**6**) through site-specific C–H oxidations of dihydrojunenol (**4**). Reagents and conditions as follows. **a**, $\text{CF}_3\text{CH}_2\text{NCO}$ (1.0 equiv.), Pyr (4.0 equiv.), DMAP (catalytic), DCM, RT, 1 h, 99%. **b**, TFDO (1.0 equiv.), DCM, -20°C , portion-wise addition of TFDO over 30 min, then additional 30 min, 82%. **c**, NaOMe (5.0 equiv.), MeOH, 70°C , 2 h, 95%. **d**, $\text{CH}_3\text{CO}_2\text{Br}$ (1.0 equiv.), DCM, 0°C , 5 min; PhCF_3 , 100-W sunlamp, 10 min; Ag_2CO_3 (1.2 equiv.),

DCM, RT, 30 min, then aqueous acetic acid, RT, 30 min; LiOH (10 equiv.), THF/ H_2O , RT, 10 min, 43% (39% recovered **15**). DMAP, 4-dimethylaminopyridine; TFDO, methyl(trifluoromethyl)dioxirane; NaOMe, sodium methoxide; THF, tetrahydrofuran. For selected physical data for compounds **5**, **6**, **15** and **16**, see the Supplementary Information. Compounds **5**, **6** and **15** were verified by X-ray crystallography.

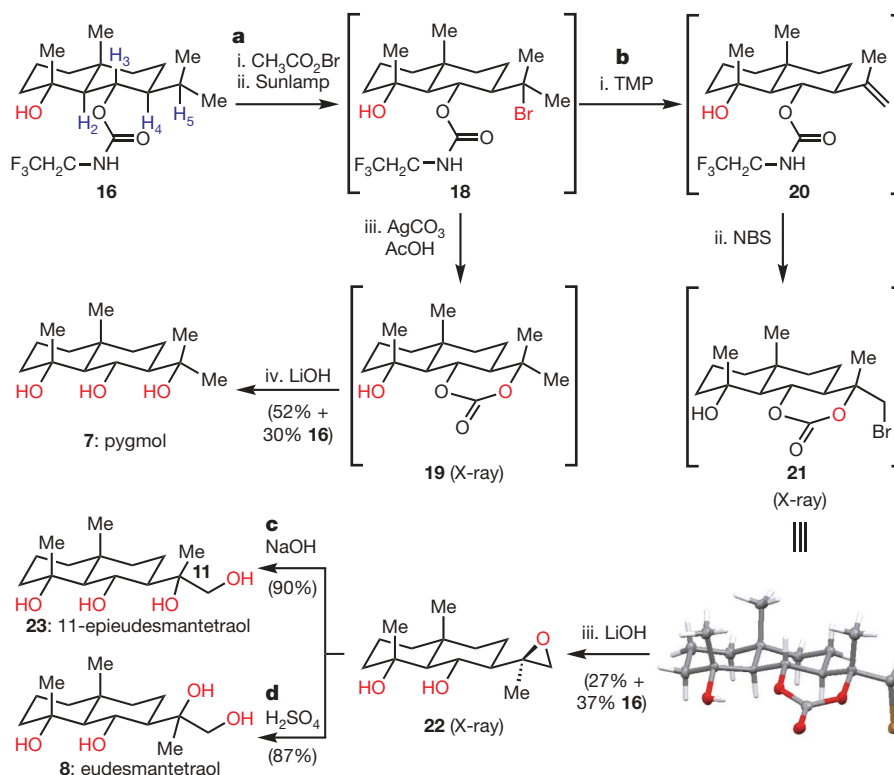


Figure 4 | Total syntheses of pygmul (**7**) and eudesmantetraol (**8**) through site-specific C–H oxidations of **16**. Reagents and conditions as follows. **a**, $\text{CH}_3\text{CO}_2\text{Br}$ (1.0 equiv.), DCM, 0°C , 5 min; PhCF_3 , 100-W sunlamp, 20 min; Ag_2CO_3 (1.2 equiv.), DCM, RT, 30 min, then aqueous acetic acid, RT, 30 min; LiOH (10 equiv.), THF/ H_2O , RT, 10 min, 52% (30% recovered **16**). **b**, TMP (2.0 equiv.), toluene, 80°C , 12 h; NBS (2.0 equiv.), DCM, RT, 6 h, then aqueous acetic acid, RT, 30 min; LiOH (10 equiv.), THF/ H_2O , RT,

10 min, 27% (37% recovered **16**). **c**, 3 M NaOH, DMSO, 80°C , 2 h, 90%. **d**, 0.1 M H_2SO_4 , DME/ H_2O , RT, 1 h, 87%. TMP, 2,2,6,6-tetramethylpiperidine; NBS, *N*-bromosuccinimide; DMSO, dimethylsulfoxide; DME, 1,2-dimethoxyethane. For selected physical data for compounds **7**, **8**, **19**, **21**, **22** and **23**, see the Supplementary Information. Compounds **19**, **21** and **22** were verified by X-ray crystallography.

neighbouring trifluoroethyl carbamate could be cyclized onto this alkene using NBS to afford the bromocarbonate **21** (structure verified by X-ray crystallography) as a single diastereomer. This two-step sequence represents a rare example of carbamate-directed remote dehydrogenation and the use of a carbamate to achieve intramolecular stereocontrolled oxybromination of an alkene. Treatment with LiOH led to epoxide intermediate **22** (structure verified by X-ray crystallography) in 27% overall isolated yield (the C–H activation step proceeds in ~60% isolated yield) for the three-step sequence. Exposure of **22** to NaOH furnished 11-epieudesmantetraol **23** with retention of stereochemistry (90% yield), whereas treatment with dilute acid afforded eudesmantetraol (**8**) as a single diastereomer in 87% yield by net inversion. Notably, the dihydroxylation of olefin **20** using OsO₄ was not stereoselective, giving a mixture of diol products (Sharpless AD-mixes were also ineffective).

As such, the terpenes cryptone (**10**), dihydrojunenol (**4**), 4-epiajanol (**5**), dihydroyeudesmane (**6**), pygmal (**7**) and eudesmantetraol (**8**) were respectively constructed in 2, 9, 12, 12, 13 and 15 steps and in 63, 21, 17, 9, 9 and 4% overall isolated yield from two feedstock carbon sources (methyl vinyl ketone and 3-methyl butyraldehyde) without any recycling of recovered starting materials. Before this work, a single late-stage C–H-to-C–OH conversion was used for the diversification of bryostatin analogues¹⁴, the semi-synthesis of various natural products such as steroids¹⁶ and artemisinin¹⁷, and the preparation of linear (E)-allylic alcohols²⁹. This work presents an example of a linear C–H activation strategy featuring multiple consecutive site-selective oxidations in total synthesis.

The straightforward logic employed in this atypical approach to terpene synthesis begins with the construction of a ‘retrosynthesis pyramid’ diagram as shown in Fig. 5. The retrosynthesis pyramid places the highest oxidized target at the apex and works backward until

the lowest oxidized members are reached. Rather than making strategic disconnections to a single molecule, multiple molecules are considered at every descending level of oxidation. Next, each individual level of ‘redox isomers’ is scrutinized for feasibility in the required oxidation to reach the next level. By analysing several molecules with equivalent overall oxidation state, there is no limitation to a certain oxidation reaction (that is, the conversion of C–H to C–X or direct dehydrogenation to an alkene can sometimes be more strategic than the conversion of C–H to C–O). It is here that issues of choreography, reaction methodology and chemoselectivity are evaluated. On reaching the lowest oxidation level, a final target is chosen for both its ease of synthesis and its potential to access the greatest number of possible intermediates as the pyramid is ascended. Because much of the chemistry necessary to reach the apex may be poorly preceded and require reaction invention, it is critical that the starting material be available in large (gram-scale) quantities. After the planning necessary to mimic the oxidase-phase has been concluded, the most logical starting material can then be evaluated using standard retrosynthetic analysis (this is the conceptual equivalent to the cyclase phase in biosynthesis). A detailed discussion of the implementation of the retrosynthesis pyramid for this work is contained in the Supplementary Information.

One of the main characteristics of terpene biosynthesis is that installation of functional groups (oxidation) onto the carbon framework usually occurs near the end of the sequence (a notable exception being the oxidation of squalene to squalene oxide^{6–8}). These syntheses demonstrate that there may be certain advantages to conducting terpene synthesis in a similar manner. Because most of the heteroatoms are installed at the end of the synthesis, chemoselectivity becomes less of a concern during the early stages²⁹. This also leads to a minimization of protecting group chemistry during the cyclase phase and allows for a greater variety of chemistry to be employed with less

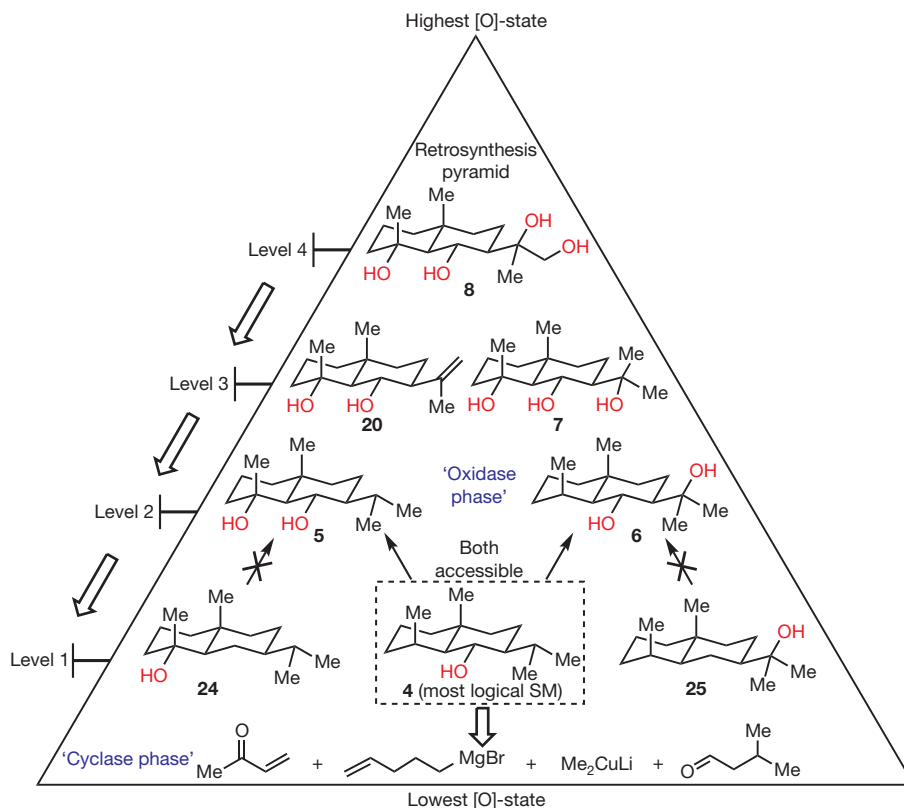


Figure 5 | Pyramid diagram for the retrosynthetic planning of terpene synthesis using a ‘two-phase’ approach. Because eudesmantetraol (**8**) is the highest oxidized target, it is placed at the apex. Removal of one hydroxyl group leads to level-3 intermediates **20** and **7** (and any synthetic equivalents such as an alkyl bromide, for example **18** in Fig. 4). Repetition of this transform leads to diols **5** and **6** (level 2), either of which could conceivably

access **20** or **7**. Subsequent deoxygenation of these level-2 intermediates leads to three selections for level 1: **24**, **4** and **25**. Dihydrojunenol (**4**) was chosen as the most logical starting material owing to its potential to access both **5** and **6** without any corrective reduction steps or a difficult C–H activation of a methylene group. [O]-state, oxidation state; SM, starting material.

concern for side reactions. Additionally, this biomimetic approach can naturally lead to the synthesis of related family members, closely related analogues (for example **5** and **23**) and in some cases even structural reassignments (for example **6'** → **6**). Finally, the use of a retrosynthesis pyramid when planning the oxidase phase can encourage the invention of useful methodology and greater insight into the relative reactivity of different tertiary C–H bonds.

This work represents another^{11–17,30} step towards the generation of a set of rules and logic for the use of C–H oxidation in terpene synthesis. Whether or not elements of such a strategy could be employed in the synthesis of terpenes at the highest level of complexity (for example **2**; Fig. 1) remains to be seen. There are obvious limitations in current synthetic methodology that will need to be overcome for the logic of this approach to reach its full potential, such as the oxidation of primary and secondary C–H bonds in a controllable manner, a broader functional group tolerance and more general, high-yielding protocols for C–H oxidation. Finally, we note that the use of a trifluoroethyl carbamate directing group is both an advantage and a limitation. Its use permitted site-selective oxidations of both *sp*³-hybridized (**15** to **17** and **16** to **18**) and *sp*²-hybridized (**20** to **21**) carbon atoms that would have been difficult to achieve in an intermolecular fashion with currently available reagents. Furthermore, it shielded the reactivity of a secondary alcohol, allowing intermolecular C–H oxidation (**15** to **16**) and rendering several intermediates crystalline for X-ray characterization purposes. Conversely, its installation requires an additional step and this points to two directions for future attempts to imitate the oxidase phase of terpene biosynthesis: site-specific C–H oxidation without any directing groups and reagent-dependent reordering of C–H bond reactivity.

METHODS SUMMARY

All reactions were carried out under a nitrogen atmosphere with dry solvents under anhydrous conditions, unless otherwise noted. Yields refer to chromatographically and spectroscopically homogeneous materials, unless otherwise stated. Reagents were purchased at the highest commercial quality and used without further purification, unless otherwise stated. Reactions were monitored using thin-layer chromatography. For full experimental details and procedures for all reactions performed and full characterization (¹H NMR, ¹³C NMR, high-resolution mass spectrometry, infrared, optical rotation, melting point and *R*_f value) of all new compounds, see Supplementary Information.

Received 17 February; accepted 7 April 2009.

Published online 13 May 2009.

- Nicolaou, K. C. & Montagnon, T. *Molecules That Changed the World* (Wiley-VCH, 2008).
- Davis, E. M. & Croteau, R. Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes and diterpenes. *Top. Curr. Chem.* **209**, 53–95 (2000).
- Maimone, T. J. & Baran, P. S. Modern synthetic efforts toward biologically active terpenes. *Nature Chem. Biol.* **3**, 396–407 (2007).
- Corey, E. J. & Cheng, X. M. *The Logic of Chemical Synthesis* (Wiley, 1995).
- Wilson, R. & Danishefsky, S. J. Pattern recognition in retrosynthetic analysis: snapshots in total synthesis. *J. Org. Chem.* **72**, 4293–4305 (2007).
- Yoder, R. A. & Johnston, J. N. A case study in biomimetic total synthesis: polyolefin carbocyclizations to terpenes and steroids. *Chem. Rev.* **105**, 4730–4756 (2005).
- Ruzicka, L. Isoprene rule and the biogenesis of terpenic compounds. *Experientia* **9**, 357–367 (1953).
- Eschenmoser, A. & Arigoni, D. Revisited after 50 years: the 'Stereochemical interpretation of the biogenetic isoprene rule for the triterpenes'. *Helv. Chim. Acta* **88**, 3011–3050 (2005).
- Jennewein, S., Rithner, C. D., Williams, R. M. & Croteau, R. B. Taxol biosynthesis: taxane 13 α -hydroxylase is a cytochrome P450-dependent monooxygenase. *Proc. Natl Acad. Sci. USA* **98**, 13595–13600 (2001).

- Denis, J. N. *et al.* Highly efficient, practical approach to natural Taxol. *J. Am. Chem. Soc.* **110**, 5917–5919 (1988).
- Mello, R., Fiorentino, M., Fusco, C. & Curci, R. Oxidations by methyl(trifluoromethyl)dioxirane. 2. Oxyfunctionalization of saturated hydrocarbons. *J. Am. Chem. Soc.* **111**, 6749–6757 (1989).
- Costas, M., Mehn, M. P., Jensen, M. P. & Que, L. Jr. Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. *Chem. Rev.* **104**, 939–986 (2004).
- Groves, J. T., Bonchio, M., Carofiglio, T., & Shalyaev, K. Rapid catalytic oxygenation of hydrocarbons by ruthenium pentafluorophenylporphyrin complexes: evidence for the involvement of a Ru(III) intermediate. *J. Am. Chem. Soc.* **118**, 8961–8962 (1996).
- Wender, P. A., Hilinski, M. K. & Mayweg, A. V. W. Late-stage intermolecular C–H activation for lead diversification: a highly chemoselective oxyfunctionalization of the C-9 position of potent bryostatin analogues. *Org. Lett.* **7**, 79–82 (2005).
- Brodsky, B. H. & Du Bois, J. Oxaziridine-mediated catalytic hydroxylation of unactivated 3° C–H bonds using hydrogen peroxide. *J. Am. Chem. Soc.* **127**, 15391–15393 (2005).
- Yang, J., Gabriele, B., Belvedere, S., Huang, Y. & Breslow, R. Catalytic oxidations of steroid substrates by artificial cytochrome P-450 enzymes. *J. Org. Chem.* **67**, 5057–5067 (2002).
- Chen, M. S. & White, M. C. A predictably selective aliphatic C–H oxidation reaction for complex molecule synthesis. *Science* **318**, 783–787 (2007).
- Wu, Q.-X., Shi, Y.-P. & Jia, Z.-J. Eudesmane sesquiterpenoids from the Asteraceae family. *Nat. Prod. Rep.* **23**, 699–734 (2006).
- Cardona, L., Garcia, B., Gimenez, E. & Pedro, J. R. A shorter route to the synthesis of (+)-junenol, isojunenol, and their coumarate esters from (–)-santonin. *Tetrahedron* **48**, 851–860 (1992).
- Levine, S. R., Krout, M. R. & Stoltz, B. M. Catalytic enantioselective approach to the eudesmane sesquiterpenoids: total synthesis of (+)-carissone. *Org. Lett.* **11**, 289–292 (2009).
- Paknikar, S. K., Dhekne, V. V. & Joshi, G. D. Synthesis of 4 α , 6 α -dihydroxyeudesmane: revision of stereochemistry at C-4 of ajanol. *Indian J. Chem.* **15B**, 86–87 (1977).
- Zhao, P.-J., Li, G.-H. & Shen, Y.-M. New chemical constituents from the endophyte *Streptomyces* species LR4612 cultivated on *Maytenus hookeri*. *Chem. Biodivers.* **3**, 337–342 (2006).
- Irwin, M. A. & Geissman, T. A. Sesquiterpene alcohols from *Artemisia pygmaea*. *Phytochemistry* **12**, 849–852 (1973).
- De Marino, S. *et al.* New sesquiterpene lactones from *Laurus nobilis* leaves as inhibitors of nitric oxide production. *Planta Med.* **71**, 706–710 (2005).
- Betancort, J. M. & Barbas, C. F. Catalytic direct asymmetric Michael reactions: taming naked aldehyde donors. *Org. Lett.* **3**, 3737–3740 (2001).
- Chi, Y. & Gellman, S. H. Diphenylprolinol methyl ether: a highly enantioselective catalyst for Michael addition of aldehydes to simple enones. *Org. Lett.* **7**, 4253–4256 (2005).
- Chen, K., Richter, J. M. & Baran, P. S. 1,3-diol synthesis via controlled, radical-mediated C–H functionalization. *J. Am. Chem. Soc.* **130**, 7247–7249 (2008).
- Schreiber, J. & Eschenmoser, A. Über die relative Geschwindigkeit der Chromosäureoxydation sekundärer, alicyclischer Alkohole. Vorläufige Mitteilung. *Helv. Chim. Acta* **38**, 1529–1536 (1955).
- Fraunhoffer, K. J., Bachovchin, D. A. & White, M. C. Hydrocarbon oxidation vs C–C bond-forming approaches for efficient syntheses of oxygenated molecules. *Org. Lett.* **7**, 223–226 (2005).
- Davies, H. M. L. & Manning, J. R. Catalytic C–H functionalization by metal carbenoid and nitrenoid insertion. *Nature* **451**, 417–424 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to A. Eschenmoser for discussions and to Bristol-Myers Squibb for financial support. M. Morón Galán and Y. Ishihara are acknowledged for technical contributions to the early stages of this project. We are grateful to B. Shi and A. Reingold for assistance with high-performance liquid chromatography and X-ray crystallographic analyses, respectively.

Author Information The X-ray crystallographic coordinates for the structures reported in this paper have been deposited at the Cambridge Crystallographic Data Centre, under deposition numbers CCDC 718278 (**5**), CCDC 719000 (**6**), CCDC 718279 (**15**), CCDC 718280 (**19**), CCDC 718281 (**21**) and CCDC 718282 (**22**). These data can be obtained free of charge from the Cambridge Crystallographic Data Centre (http://www.ccdc.cam.ac.uk/data_request/cif). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P. S. B. (pbaran@scripps.edu).

The proportionality of global warming to cumulative carbon emissions

H. Damon Matthews¹, Nathan P. Gillett², Peter A. Stott³ & Kirsten Zickfeld²

The global temperature response to increasing atmospheric CO₂ is often quantified by metrics such as equilibrium climate sensitivity and transient climate response¹. These approaches, however, do not account for carbon cycle feedbacks and therefore do not fully represent the net response of the Earth system to anthropogenic CO₂ emissions. Climate–carbon modelling experiments have shown that: (1) the warming per unit CO₂ emitted does not depend on the background CO₂ concentration²; (2) the total allowable emissions for climate stabilization do not depend on the timing of those emissions^{3–5}; and (3) the temperature response to a pulse of CO₂ is approximately constant on timescales of decades to centuries^{3,6–8}. Here we generalize these results and show that the carbon–climate response (CCR), defined as the ratio of temperature change to cumulative carbon emissions, is approximately independent of both the atmospheric CO₂ concentration and its rate of change on these timescales. From observational constraints, we estimate CCR to be in the range 1.0–2.1 °C per trillion tonnes of carbon (Tt C) emitted (5th to 95th percentiles), consistent with twenty-first-century CCR values simulated by climate–carbon models. Uncertainty in land-use CO₂ emissions and aerosol forcing, however, means that higher observationally constrained values cannot be excluded. The CCR, when evaluated from climate–carbon models under idealized conditions, represents a simple yet robust metric for comparing models, which aggregates both climate feedbacks and carbon cycle feedbacks. CCR is also likely to be a useful concept for climate change mitigation and policy; by combining the uncertainties associated with climate sensitivity, carbon sinks and climate–carbon feedbacks into a single quantity, the CCR allows CO₂-induced global mean temperature change to be inferred directly from cumulative carbon emissions.

We propose a new measure of the climate response to anthropogenic carbon dioxide emissions: the ‘carbon–climate response’

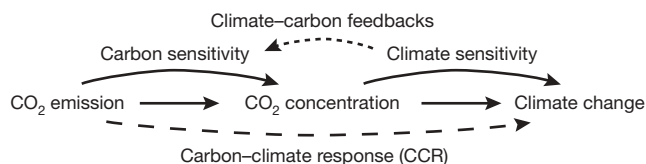


Figure 1 | Schematic representation of the progression from CO₂ emissions to climate change. We define ‘carbon sensitivity’ as the increase in atmospheric CO₂ concentrations that results from CO₂ emissions, as determined by the strength of natural carbon sinks. ‘Climate sensitivity’ is shown here as a general characterization of the temperature response to atmospheric CO₂ changes. Feedbacks between climate change and the strength of carbon sinks are shown as the upper dotted arrow (climate–carbon feedbacks). The CCR aggregates the climate and carbon sensitivities (including climate–carbon feedbacks) into a single metric representing the net temperature change per unit carbon emitted.

(CCR). The CCR is illustrated schematically in Fig. 1, which shows the progression from carbon emissions to climate change. The CCR incorporates the standard concept of climate sensitivity (the temperature response to increased atmospheric CO₂), in addition to a ‘carbon sensitivity’ (the amount by which atmospheric CO₂ concentrations increase in response to CO₂ emissions, as mediated by natural carbon sinks, and including also the effect of feedbacks between climate change and carbon uptake).

The CCR thus represents the net climate response to CO₂ emissions, and can be defined as $\Delta T/E_T$, where ΔT is the global mean temperature change over some period of time, and E_T is the total cumulative carbon dioxide emitted over that period. We assign units of trillion tonnes of carbon to E_T (1 Tt = 1 teratonne, or 10^{18} grams, of carbon, which is equivalent to 3.7 trillion tonnes of CO₂), so the CCR as defined here carries units of °C per Tt C emitted. CCR can be written as:

$$\text{CCR} = \Delta T/E_T \\ = (\Delta T/\Delta C_A) \times (\Delta C_A/E_T)$$

where ΔC_A is the change in atmospheric carbon (in Tt C). Written in this way, CCR represents the product of the temperature change per unit atmospheric carbon increase ($\Delta T/\Delta C_A$) and the airborne fraction of cumulative carbon emissions ($\Delta C_A/E_T$). If defined under conditions of constant doubled pre-industrial atmospheric CO₂, ΔT is equal to the equilibrium climate sensitivity, and if defined under doubled CO₂ conditions in a simulation in which CO₂ increases at 1% per year, ΔT is equal to the transient climate response¹.

Both the airborne fraction of cumulative emissions and the temperature change per unit atmospheric carbon increase are dependent on the atmospheric CO₂ concentration and its rate of increase; however, the CCR (as the product of the two) shows a remarkable constancy with time. This can be seen in Fig. 2, which shows three model simulations using the University of Victoria Earth System Climate Model⁹ (UVic ESCM, see Methods), an intermediate-complexity coupled climate–carbon model. In all simulations, we prescribed atmospheric CO₂ concentrations and used the model’s interactive carbon sinks to diagnose the implied anthropogenic CO₂ emissions consistent with the prescribed concentration changes¹⁰. In the first simulation (Fig. 2a) we increased atmospheric CO₂ by 1% per year for 70 years; in the second and third simulations (Fig. 2b), atmospheric CO₂ was doubled (solid lines) or quadrupled (dashed lines) instantaneously and held constant for 1,000 years. In all simulations, the airborne fraction of cumulative emissions decreased over time, whereas the temperature change per unit change in atmospheric carbon increased with time. After an initial adjustment period of about a decade, the CCR remained almost constant at ~ 1.7 °C per Tt C emitted.

¹Department of Geography, Planning and Environment, Concordia University, 1455 de Maisonneuve Blvd W., Montreal, Quebec, H3G 1M8, Canada. ²Canadian Centre for Climate Modelling and Analysis, Environment Canada, 3800 Finnerty Road, Victoria, British Columbia, V8P 5C2, Canada. ³Met Office Hadley Centre, FitzRoy Road, Exeter, Devon, EX1 3PB, UK.

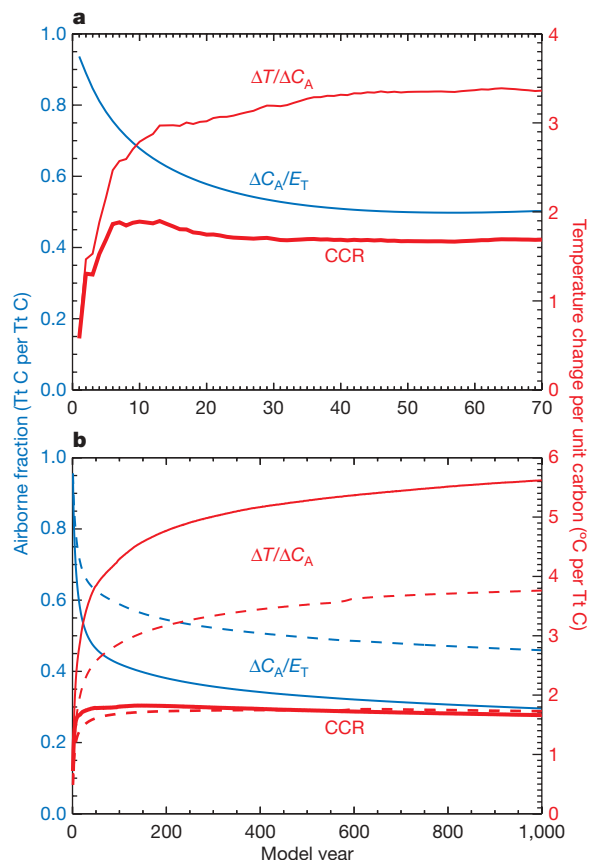


Figure 2 | Idealized model simulations of the CCR. **a**, Simulation with a 1% per year atmospheric CO_2 increase for 70 years, showing temperature change per unit atmospheric carbon increase ($\Delta T/\Delta C_A$; thin red line, right axis), airborne fraction of cumulative carbon emissions ($\Delta C_A/E_T$; thin blue line, left axis) and CCR (thick red line, right axis). In this simulation, cumulative airborne fraction decreased with time owing to a delayed carbon cycle response to a rapid prescribed rate of atmospheric CO_2 increase. This is consistent with saturating carbon sinks at higher atmospheric CO_2 , which leads to an increased airborne fraction of annual emissions with increasing atmospheric CO_2 . **b**, Simulations with an instantaneous doubling (solid lines) and quadrupling (dashed lines) of atmospheric CO_2 for 1,000 years (colours as in **a**). In all cases, the cumulative airborne fraction decreased with time, whereas the temperature change per unit atmospheric carbon increased with time; consequently, the CCR (defined as the product of these two quantities) remained constant in time.

In these simulations, the CCR is independent of both time and CO_2 emission (or concentration) scenario. At a given CO_2 concentration (see, for example, Fig. 2b), the time-independence of CCR arises from a cancellation of a decreasing airborne fraction of cumulative emissions, and an increasing temperature change per unit atmospheric CO_2 over time. This may relate in part to the uptake of heat and carbon by the ocean being driven by the same deep-ocean mixing processes on long timescales^{3,7}. However, as can be seen in Fig. 2a and b, CCR is also independent of CO_2 concentration and, by extension, of the CO_2 emission scenario. This scenario independence emerges owing to the approximate cancellation of the saturation of carbon sinks and the saturation of CO_2 radiative forcing with increasing atmospheric CO_2 . As a result, at higher atmospheric concentrations, a given CO_2 emission will lead to a larger increase in atmospheric CO_2 , but the temperature change per unit change in atmospheric CO_2 will be smaller².

Even in the extreme case of instantaneous pulse emissions⁸, the temperature change per unit carbon emitted in the UVic ESCM is found to be constant to within 10% on timescales of between 20 and 1,000 years, and for cumulative emissions of up to 2 Tt C (see Supplementary Fig. 1). As is seen, however, in Fig. 2a, we expect

CCR to be more closely constrained in simulations in which cumulative emissions vary smoothly. Nonetheless, if used as a metric for model intercomparison, we recommend that CCR be defined under standard conditions, such as at the time of CO_2 doubling in a transient simulation with 1% CO_2 increase per year. Defined in this way, CCR generalizes previously proposed metrics (such as the temperature response to a small pulse or constant sustained emission⁶—see Supplementary Information for additional discussion) into a single robust and versatile quantity which can be easily estimated from current standard model experiments, and yet represents the climate response to a wide range of CO_2 emissions scenarios.

In a given model, CCR is approximately constant with respect to time and emissions scenario; however, we would expect the value of CCR to vary among models owing to differences in both climate and carbon sensitivities. Its time and scenario independence mean that the CCR can be estimated from any model simulation with either prescribed CO_2 emissions, or prescribed CO_2 concentrations and prognostic model carbon sinks. Consequently, the simulations performed as part of the Coupled Climate Carbon Cycle Model Intercomparison Project (C4MIP¹¹) provide a means of estimating the range of CCR values among the current generation of coupled climate–carbon models.

Figure 3 shows results from the 11 C4MIP models and the ensemble mean, with global temperature change plotted as a function of cumulative carbon emissions (Fig. 3a) and temperature change per unit carbon emitted plotted as a function of time (Fig. 3b). Most models (and the ensemble mean) show a nearly linear relationship between temperature change and cumulative emissions (Fig. 3a), suggesting that this may be a robust property of the coupled climate–carbon system. Some models do deviate from linearity, particularly early in the simulations, which is at least partly due to the influence of decadal temperature variability. However, by the middle of the twenty-first century, all models converge to an intrinsic value of temperature change per unit carbon emitted, which remains approximately stable for the remainder of the simulation (Fig. 3b). CCR values calculated at the time of CO_2 doubling in each model simulation are given in Supplementary Table 1. Model values of CCR range from 1.0 to 2.1 $^\circ\text{C}$ per Tt C, with an ensemble mean value of 1.6 $^\circ\text{C}$ per Tt C (see Supplementary Information for additional discussion of model CCR values).

The CCR can also be estimated from historical carbon emissions data and observed temperature changes. To calculate CCR from observations, we first estimated decadal-mean CO_2 -attributable warming relative to 1900–09 by scaling an estimate of greenhouse-gas-attributable warming¹² by the ratio of CO_2 to greenhouse-gas forcing. We then calculated CCR by dividing CO_2 -attributable warming by cumulative anthropogenic CO_2 emissions between 1900–09 and each subsequent decade, including emissions from land-use change, fossil fuels and cement production (see Methods).

Figure 4 shows an estimate of CCR for 1990–99 of 1.0–2.1 $^\circ\text{C}$ per Tt C (5 to 95% confidence interval), with a best estimate of 1.5 $^\circ\text{C}$ per Tt C. Similar estimates of CCR, albeit with larger uncertainties, are obtained for previous decades. We note that these estimates are less contaminated with internal climate variability than those derived from single simulations in Fig. 3 because the greenhouse-gas-attributable warming is based on a scaled ensemble mean of 11 simulations. Nonetheless, assuming the simulated temporal evolution of the greenhouse gas response is realistic, these results provide further evidence for the constancy of CCR as a function of time.

Recent climate–carbon model experiments have shown that eliminating CO_2 emissions leads to approximately stable, or slowly decreasing global temperatures over time^{3,7,13}; this implies that close to zero net anthropogenic carbon emissions are required to stabilize global mean temperature³, and conversely that there may be negligible future warming commitment as a result of past CO_2 emissions^{3,7,13}. Consequently, the CCR, defined here as the ratio of instantaneous temperature change to past CO_2 emissions, can also

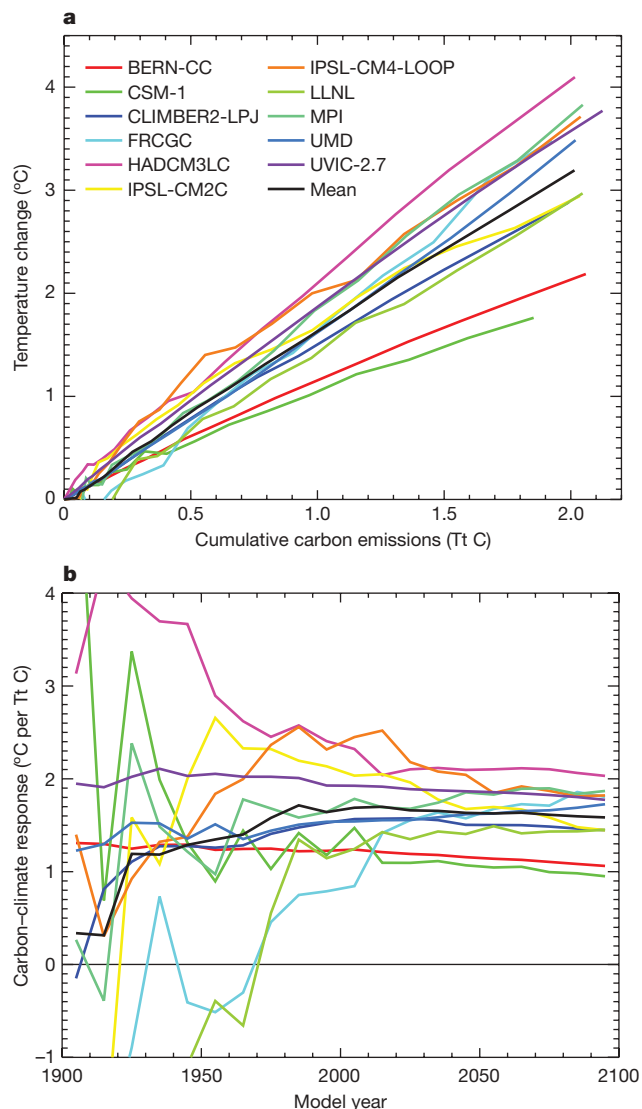


Figure 3 | CCR estimated from the C4MIP simulations¹¹. **a**, Decadal-average temperature change plotted as a function of cumulative carbon emissions, showing a near-linear relationship for both individual models (coloured lines) and the ensemble mean (black line). **b**, Temperature change per cumulative carbon emitted for each decade from 1900 to 2100 relative to the first decade of each model simulation. Over most of the twenty-first century portion of the simulations, CCR values in each model are remarkably constant in time.

be used as an estimate of the centennial-scale temperature legacy of these emissions. As a result, our estimates of CCR can be inverted to estimate the total allowable anthropogenic carbon emissions per degree of long-term temperature change.

From our model-based estimate of CCR, we estimate allowable emissions of 1.25 Tt C (range, 0.95–2 Tt C) for 2 °C warming relative to pre-industrial temperature; our observationally based best estimate of allowable emissions for 2 °C of warming is 1.4 Tt C (5–95% confidence interval, 1.0 to 1.9 Tt C). Given total CO₂ emissions until now of approximately 0.5 Tt C from fossil fuels and land-use change^{14,15}, this implies that total future carbon emissions consistent with 2 °C of warming must be restricted to a best estimate of about 0.8 Tt C (0.7 Tt C based on the model ensemble mean; 0.9 Tt C based on observational constraints).

We emphasize, however, that the calculated uncertainty on this number is quite large (0.4 to 1.5 Tt C). Furthermore, we are unable to exclude the possibility of higher values of CCR (and consequently lower values of allowable emissions), owing particularly to poorly

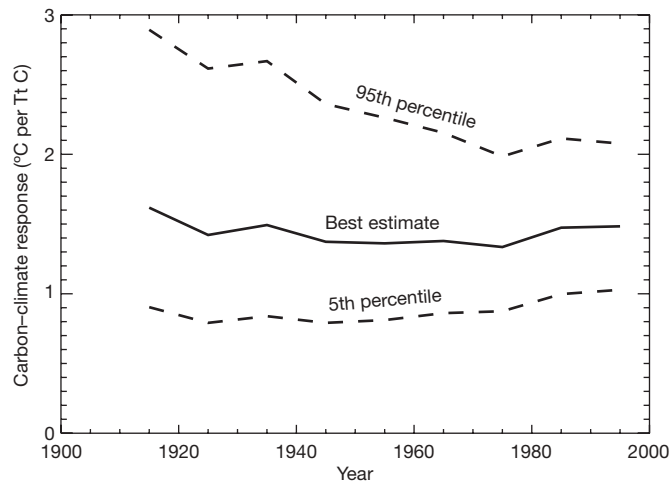


Figure 4 | Observational estimates of CCR. CCR was estimated for each decade of the twentieth century after 1910 by scaling an observationally constrained estimate of greenhouse-gas-attributable warming relative to 1900–09 by the ratio of CO₂ forcing to total greenhouse gas forcing, and dividing by cumulative anthropogenic carbon emissions over the same period. This observationally constrained estimate of CCR is both stable in time and consistent with the estimates derived from model simulations.

quantified uncertainties in historical land-use change emissions and structural uncertainties in the simulated sulphate aerosol response. For example, the allowable emissions for a particular warming target calculated by ref. 5 were lower, because they used a higher observational estimate of CO₂-attributable warming as well as a climate–carbon model which simulated non-negligible zero emissions commitment under conditions of high climate sensitivity. We note also that our analysis of allowable emissions applies specifically to CO₂-induced warming, and does not account for the effects of other greenhouse gases or aerosols.

The CCR is a simple, yet robust representation of the global temperature response to anthropogenic CO₂ emissions, and as such is directly relevant to current policy negotiations surrounding international climate mitigation efforts. The European Union has proposed restricting global warming to less than 2 °C above pre-industrial temperatures¹⁶; however, large uncertainty in equilibrium climate sensitivity¹⁷ prevents confident estimates of the CO₂ stabilization level required to avoid 2 °C warming, and climate sensitivity alone provides no policy-useful information about the allowable CO₂ emissions for a given stabilization level. The CCR represents a synthesis of previous efforts to quantify the temperature response to anthropogenic CO₂ emissions by aggregating the uncertainties associated with climate sensitivity, carbon sinks and climate–carbon feedbacks into a single well-constrained metric of climate change that is related directly to cumulative carbon emissions.

METHODS SUMMARY

For the idealized model experiments (1% per year CO₂ increase; doubled/quadrupled CO₂) we used the UVic ESCM version 2.8 (refs 9, 18–20). The UVic ESCM is a computationally efficient coupled climate–carbon model, with interactive representations of three-dimensional ocean circulation, atmospheric energy and moisture balances, sea ice dynamics and thermodynamics, dynamic vegetation and the global carbon cycle (including land and both inorganic and organic ocean carbon). Version 2.7 of the UVic ESCM was one of the 11 participating models in C4MIP¹¹, in which models were driven by a common CO₂ emissions scenario and carbon sinks and atmospheric CO₂ concentrations were calculated interactively until the year 2100. From the C4MIP simulations, we estimated CCR using globally averaged temperature change and accumulated carbon emissions at the year of CO₂ doubling in each simulation.

Our observational estimate of CCR was derived using estimates of CO₂-attributable warming and cumulative CO₂ emissions for each decade of the twentieth century relative to 1900–09. We estimated CO₂-attributable warming using an estimate of greenhouse-gas-attributable warming¹², scaled by the ratio of CO₂ to

total greenhouse-gas forcing²¹, where greenhouse-gas forcing was first scaled by an estimate of the mean efficacy of long-lived greenhouse gases²². We calculated uncertainties in greenhouse-gas-attributable warming, accounting for internal variability and inter-model uncertainty¹², and assumed normally and Student-*t* distributed uncertainties for radiative forcings and greenhouse-gas efficacy, respectively²². We calculated cumulative carbon emissions from fossil fuels and land-use change^{13,14,23}, and assumed a one-sigma systematic uncertainty on land-use emissions of ± 0.5 Pg C per year²⁴. Our central estimates for CO₂-attributable warming and cumulative emissions at 1990–99 relative to 1900–09 were 0.492 °C and 0.338 Tt C, respectively. We calculated a probability density function for CCR based on the probability distributions of the constituent terms, which we used to estimate the mean and the 5th and 95th percentiles.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 December 2008; accepted 14 April 2009.

1. Randall, D. A. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 589–845 (Cambridge University Press, 2007).
2. Caldeira, K. & Kasting, J. F. Insensitivity of global warming potentials to carbon dioxide emissions scenarios. *Nature* **366**, 251–253 (1993).
3. Matthews, H. D. & Caldeira, K. Stabilizing climate requires near-zero emissions. *Geophys. Res. Lett.* **35**, L04705 (2008).
4. Zickfeld, K., Eby, M., Matthews, H. D. & Weaver, A. J. Setting cumulative emissions targets to reduce the risk of dangerous climate change. *Proc. Natl Acad. Sci. USA*. (submitted).
5. Allen, M. R. *et al.* Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* **458**, 1163–1166 (2009).
6. Shine, K. P., Fuglestad, J. S., Haillemariam, K. & Struber, N. Alternatives to the global warming potential for comparing climate impacts of emissions of greenhouse gases. *Clim. Change* **68**, 281–302 (2005).
7. Solomon, S., Kasper Plattner, G., Knutti, R. & Friedlingstein, P. Irreversible climate change due to carbon dioxide emissions. *Proc. Natl Acad. Sci. USA* **106**, 1704–1709 (2009).
8. Eby, M. *et al.* Lifetime of anthropogenic climate change: millennial time-scales of potential CO₂ and surface temperature perturbations. *J. Clim.* **22**, 2501–2511 (2009).
9. Weaver, A. J. *et al.* The UVic Earth System Climate Model: model description, climatology and applications to past, present and future climates. *Atmos. Ocean* **39**, 361–428 (2001).
10. Matthews, H. D. Emissions targets for CO₂ stabilization as modified by carbon cycle feedbacks. *Tellus* **55B**, 591–602 (2006).
11. Friedlingstein, P. *et al.* Climate-carbon cycle feedback analysis, results from the C4MIP model intercomparison. *J. Clim.* **19**, 3337–3353 (2006).
12. Huntingford, C., Stott, P. A., Allen, M. R. & Lambert, F. H. Incorporating model uncertainty into attribution of observed temperature change. *Geophys. Res. Lett.* **33**, L05710 (2006).
13. Plattner, G.-K. *et al.* Long-term climate commitments projected with climate-carbon cycle models. *J. Clim.* **21**, 2721–2751 (2008).
14. Marland, G., Boden, T. A. & Andres, R. J. Global, regional, and national fossil fuel CO₂ emissions. In *TRENDS, A Compendium of Data on Global Change* (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US DOE, 2008).
15. Houghton, R. A. Carbon flux to the atmosphere from land-use changes: 1850–2005. In *TRENDS, A Compendium of Data on Global Change* (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US DOE, 2008).
16. European Commission. Limiting global climate change to 2 degrees Celsius: the way ahead for 2020 and beyond. (Commission of the European Communities, 2007).
17. Meehl, G. A. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 747–845 (Cambridge University Press, 2007).
18. Meissner, K. J., Weaver, A. J., Matthews, H. D. & Cox, P. M. The role of land-surface dynamics in glacial inception: a study with the UVic Earth System Climate Model. *Clim. Dyn.* **21**, 515–537 (2003).
19. Matthews, H. D., Weaver, A. J. & Meissner, K. J. Terrestrial carbon cycle dynamics under recent and future climate change. *J. Clim.* **18**, 1609–1628 (2005).
20. Schmittner, A., Oschlies, A., Matthews, H. D. & Galbraith, E. D. Future changes in climate, ocean circulation, ecosystems and biogeochemical cycling simulated for a business-as-usual CO₂ emissions scenario until year 4000 AD. *Glob. Biogeochem. Cycles* **22**, GB1013 (2008).
21. Gregory, J. M. & Forster, P. M. Transient climate response estimated from radiative forcing and observed temperature change. *J. Geophys. Res.* **113**, D23105 (2008).
22. Forster, P. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 129–234 (Cambridge University Press, 2007).
23. Denman, K. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 129–234 (Cambridge University Press, 2007).
24. Canadell, J. G. *et al.* Contributions to accelerating atmospheric CO₂ growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proc. Natl Acad. Sci. USA* **104**, 18866–18870 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Weaver, M. Eby, V. Arora, N. Ramankutty, M. Allen, S. Solomon, K. Keller, K. Caldeira and S. Turner for commentary and discussions on this work. We also thank P. Forster for providing radiative forcing time series, and P. Friedlingstein and the C4MIP modelling community for the availability of their model output. H.D.M. acknowledges support from the National Science and Engineering Research Council of Canada, and the Canadian Foundation for Climate and Atmospheric Sciences Project Grants. P.A.S. was supported by the Joint DECC, Defra and MoD Integrated Climate Programme. N.P.G. received support from the Leverhulme Trust. N.P.G. and P.A.S. acknowledge support from the Climate Change Detection and Attribution Project, jointly funded by NOAA's Office of Global Programs and the US Department of Energy.

Author Contributions H.D.M. proposed the study, carried out model simulations and analysis, and wrote most of the paper. N.P.G. proposed the inclusion of observational constraints, N.P.G. and P.A.S. carried out this analysis, and N.P.G. wrote the sections of the paper and methods describing these results. K.Z. provided additional model simulations and analysis as described in the Supplementary Information. All authors participated in discussions pertaining to interpretation and presentation of results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.D.M. (dmatthew@alcor.concordia.ca).

METHODS

UVic ESCM. The UVic ESCM is an intermediate-complexity coupled climate-carbon model. The climate component consists of a reduced-complexity energy-moisture balance atmosphere coupled to a general circulation ocean and dynamic/thermodynamic sea-ice model⁹. The carbon cycle component of version 2.8 consists of a biochemical dynamic vegetation model^{18,19} and an organic/inorganic ocean carbon cycle model²⁰. Version 2.7 of the UVic ESCM was one of the 11 participating models in the C4MIP¹¹, as well as a contributing model to the long-term climate and carbon cycle projections highlighted in ref. 17. **C4MIP.** The C4MIP compared the simulated climate and carbon cycle changes from 11 coupled climate-carbon models (including seven atmosphere-ocean general circulation models, and four intermediate-complexity models)¹¹. Models were driven by a common CO₂ emissions scenario (including specified emissions from both fossil fuels and land-use change), with carbon sinks and atmospheric CO₂ calculated interactively until the year 2100. To calculate the CCR for each model, we used globally averaged temperature changes from the coupled simulations, along with a running total of specified CO₂ emissions. The values of CCR presented here and in the Supplementary Information were calculated using a ten-year average of temperature increases and cumulative emissions, centred at the time of CO₂ doubling in each simulation.

Observationally constrained CCR estimate. We calculated observational estimates of CCR by taking the ratio of CO₂-attributable warming and cumulative emissions in the decade 1900–09 and each subsequent decade of the twentieth century. We began with a multi-model estimate of greenhouse-gas-attributable warming for each decade of the twentieth century. This was derived by scaling the mean simulated temperature response to prescribed historical well-mixed greenhouse-gas concentrations from HadCM3, GFDL and PCM to best-fit HadCRUT2v temperature observations, based on a multiple regression together with the response to sulphate aerosol and natural forcing¹². The calculated uncertainty in this greenhouse-gas-attributable warming includes an estimate of internal variability based on control simulations and an estimate of model uncertainty based on inter-model differences in forcings and simulated responses¹².

We scaled the greenhouse-gas-attributable warming by the ratio of CO₂ forcing to total well-mixed greenhouse gas forcing, with all forcings expressed as differences between 1900–09 and subsequent decades of the twentieth century²¹. Before this scaling, we multiplied the well-mixed greenhouse-gas forcing by the mean

efficacy for long-lived greenhouse gases (shown in figure 2.19 of ref. 22) to account for the larger temperature response per unit radiative forcing for other greenhouse gases compared to CO₂. Tropospheric ozone changes were not specified in the simulations used by ref. 12, so we did not include them in our estimate of total greenhouse gas forcing, under the assumption that the response to tropospheric ozone is spatially and temporally dissimilar to that due to the well-mixed greenhouse gases and is therefore unlikely to be aliased in the multiple regression (the inclusion of tropospheric ozone forcing in the total greenhouse-gas forcing estimate reduces our observational estimate of CCR to 0.9–1.8 °C per Tt C). Our calculation also assumes that climate forcings other than CO₂ emissions have had little influence on atmospheric CO₂ concentration. This is a reasonable assumption given a near-cancellation over the past century of positive non-CO₂ greenhouse-gas forcing and negative aerosol forcing.

Uncertainties in greenhouse-gas-attributable warming were calculated following ref. 12; uncertainties in radiative forcings were estimated from ref. 22 (FAQ 2.1, Fig. 2) and were assumed to be normally distributed; uncertainties in efficacies were estimated from figure 2.19 of ref. 22, and were assumed to be Student-*t* distributed. Land use, fossil fuel and cement emissions were taken from CDIAC^{14,15}. A one-sigma uncertainty on fossil fuel and cement emissions of ±5% was assumed following ref. 23 and a one-sigma systematic uncertainty on land-use emissions of ±0.5 Pg C per year was assumed following ref. 24; both were assumed to be normally distributed. A probability density function was calculated for CCR based on the probability density functions of the constituent terms, and this was used to derive the mean and the 5th and 95th percentiles. The uncertainty in land-use emissions was the largest single contributor to the overall uncertainty in CCR. Given this, we tested the sensitivity of our results to setting land-use emissions to zero; this gave an estimate of CCR for the decade 1990–99 of 1.6–2.7 °C per Tt C, though we emphasize that this should not be taken as a realistic upper bound for CCR, because zero land-use emissions are not consistent with observed atmospheric CO₂ increases. Uncertainties in the overall magnitude of aerosol forcing are fully accounted for in our estimate of greenhouse-gas-attributable warming; however, uncertainties in the temporal or spatial pattern of the response to aerosol forcing are only accounted for to the extent that they are sampled in the three global climate models we used, and errors in these patterns could lead to values of CCR outside our estimated uncertainty range.

Slow earthquakes triggered by typhoons

ChiChing Liu¹, Alan T. Linde² & I. Selwyn Sacks²

The first reports^{1,2} on a slow earthquake were for an event in the Izu peninsula, Japan, on an intraplate, seismically active fault. Since then, many slow earthquakes have been detected^{3–8}. It has been suggested⁹ that the slow events may trigger ordinary earthquakes (in a context supported by numerical modelling¹⁰), but their broader significance in terms of earthquake occurrence remains unclear. Triggering of earthquakes has received much attention: strain diffusion from large regional earthquakes has been shown to influence large earthquake activity^{11,12}, and earthquakes may be triggered during the passage of teleseismic waves¹³, a phenomenon now recognized as being common^{14–17}. Here we show that, in eastern Taiwan, slow earthquakes can be triggered by typhoons. We model the largest of these earthquakes as repeated episodes of slow slip on a reverse fault just under land and dipping to the west; the characteristics of all events are sufficiently similar that they can be modelled with minor variations of the model parameters. Lower pressure results in a very small unclamping of the fault that must be close to the failure condition for the typhoon to act as a trigger. This area experiences very high compressional deformation but has a paucity of large earthquakes; repeating slow events may be segmenting the stressed area and thus inhibiting large earthquakes, which require a long, continuous seismic rupture.

Taiwan is located along the boundary between the Philippine Sea plate and the Eurasian plate, one of the most active plate boundaries in the world (Fig. 1). The oblique collision between these two plates drives the mountain building¹⁸ and high seismic activity¹⁹ in this area. Taiwan can be divided into three major geologic provinces: the Central Range, the Western Foothills and the Coastal Range²⁰. The Central Range forms the backbone ridge of the island, where the older continental shelf and slope sediments are raised to a maximum elevation of almost 4,000 m by the on-going orogenic process. The Western Foothills, composed of Oligocene to Pleistocene clastic sediments in a fold-and-thrust belt, suffered a destructive earthquake, of moment magnitude $M_w = 7.7$, on 21 September 1999. The third province is the Coastal Range in eastern Taiwan. Neogene rocks in this province were folded into a series of north-northeast-trending anticlines and synclines and have been cut by thrust faults mostly younger than the early Pleistocene epoch²¹.

The uplift rates along the coast, derived both from geodetic levelling for the past two decades²² and from marine terraces for the past 15,000 years²³, show a consistent pattern with uplift of ~ 10 – 20 mm yr^{-1} at the southern end, decreasing to ~ 0 mm yr^{-1} at the northern end. Along the central section of the western boundary of this province, a creeping zone about 20 km in length (Fig. 1), monitored with levelling²⁴ and creepmeters²⁵, showed uplift and thrust shortening rates both of 20 mm yr^{-1} . The creep data are characterized by numerous episodic creep events. In the past 100 years, the Coastal Range has had only two earthquakes with local magnitudes $M_L \geq 7$ (Fig. 1) ($M_L = 7.0$ in 1937 and $M_L = 7.3$ in 1951, with surface rupture displacement of ~ 2 m) and 13 $M_L = 6.0$ – 6.7 events;

earthquakes release only a small fraction of the deformation, which accumulates at the rate of 8.2 cm yr^{-1} .

A small network (~ 10 -km aperture) of three borehole strainmeters (dilatometers) of the Sacks–Evertson design²⁶ was installed at the northern end of the creeping zone (Fig. 1, upper inset), beginning in 2002, as part of an integrated geodetic network. They were installed at depths of 200–266 m. Data from all sites exhibit characteristics of good-quality borehole strain data with response to seismic waves, clear solid-earth tidal signals and exponentially slowing contractional strain rates as the site recovers from the hole-drilling disturbance.

Taiwan experiences typhoons frequently, mainly during the second half of the calendar year. The associated lowering of air pressure normally results in a corresponding episode of expansion (positive strain) in the strainmeter data, with the amplitude of the excursion being proportional to the pressure decrease and also dependent on the bulk modulus of the surrounding rock (Fig. 2, inset). Unexpectedly, however, we also observe slow negative strain changes (contraction) coinciding with some typhoons; the largest such event

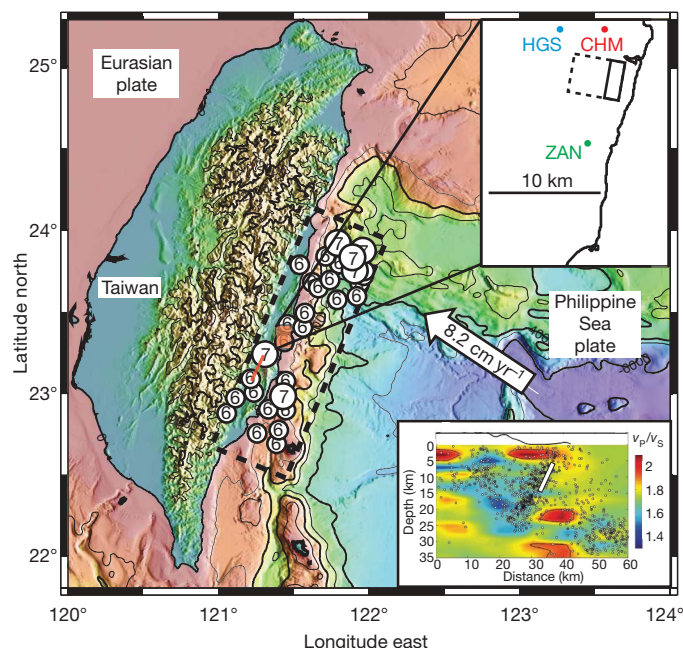


Figure 1 | Topographic map of Taiwan showing collision with Philippine Sea plate. Earthquakes shown (in dashed rectangular area) are all since 1900 and have $M_L > 6$. Of these only one (in 1972) could possibly relieve slip on our model source. Red thick line indicates a creeping zone. Upper inset: study area, showing site locations and plan view of model rupture surface for slow earthquake. Lower inset: seismic section for a profile close to the sites, indicative of a west-dipping normal fault. White rectangle shows location of model rupture. v_p , P-wave velocity; v_s , S-wave velocity.

¹Institute for Earth Sciences, Academia Sinica, 128 Sinica Road, Sec. 2, Nankang, Taipei 115, Taiwan, China. ²Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, NW, Washington DC 20015, USA.

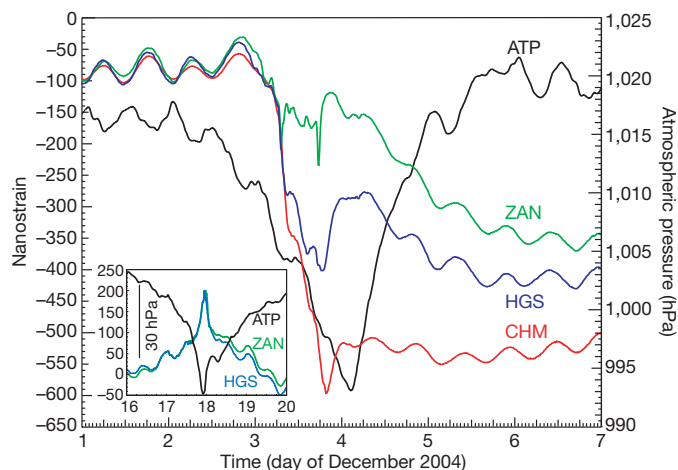


Figure 2 | Strain data for the largest slow event recorded, together with pressure record showing passage of typhoon. Coherent tidal signals are clearly visible. ZAN and HGS, with lower-modulus rock, show greater sensitivity to pressure changes than does CHM, as can be seen in strain rates from 4 to 6 December after the slow earthquake; removal of the pressure effect (Fig. 3) eliminates those trends. All sites experience negative strain changes with some subsequent recovery. This contrasts with the expected behaviour (inset, where the horizontal shows time as day of July 2005), namely that the strain should show expansion followed by recovery due to change of pressure on the surface.

is shown in Fig. 2. From a survey of the strain data (from late 2002 to late 2007), we have identified approximately 20 such strain excursions, all with net contraction (negative strain). These events have durations ranging from hours to more than a day and are detected at all sites recording data at the time of the event. Eleven events, and all those with larger amplitudes and more complex waveform character, occur during the passage of a typhoon. We do not detect any slow events during the typhoon-free first four months of the year. The

probability that 11 of these slow strain events coincide with typhoons (28 during our data coverage) by chance is vanishingly small ($<10^{-6}$). (See Supplementary Information for details of this and for additional examples of slow earthquake recordings.)

If these changes were due to surface phenomena associated with the typhoons, we would expect to see them in association with all typhoons; however, most generate only the expected expansion due to lowering of surface pressure (Fig. 2, inset). About half of the events are not associated with typhoon activity. Rather, the slow events must be due to a subsurface source and, given the tectonic setting, the most likely candidate is fault slip with rise time and propagation velocity much less than those for ordinary earthquakes—that is, a slow earthquake.

Original data for the largest observed slow event are shown in Fig. 2; in Fig. 3 we show data for the same event after removing tidal frequencies and strain changes due to atmospheric pressure changes. Sites HGS and ZAN are more influenced by pressure changes than is CHM; that is expected because of the much lower values for bulk modulus at those two sites. Other slow events are similar in character but have smaller amplitudes and are often shorter in duration (corresponding to just the first subevent; see below). The three records have much in common but also have characteristic differences. All sites have initial contraction but ZAN shows a reversal early in the record; all show reversal at a later time. Both the reversals and their differing times are indicative of a propagating source such that a nodal line in the surface strain pattern propagates past the stations and does so at different times.

Three sites do not allow an independent determination of source parameters, so we use local seismicity²⁷ (Fig. 1, lower inset) as a guide for possible source faults. These data indicate, close to our strain-meter net, a subsurface reverse fault dipping to the west with its projected intersection with the surface close to the coastline. In other cross-sections, the data indicate an east-dipping reverse fault farther to the west; we do not show it here, but slip on such a fault can satisfy the data equally well. We assume the west-dipping fault to be a source because this geometry is ideal for increasing the Coulomb criterion

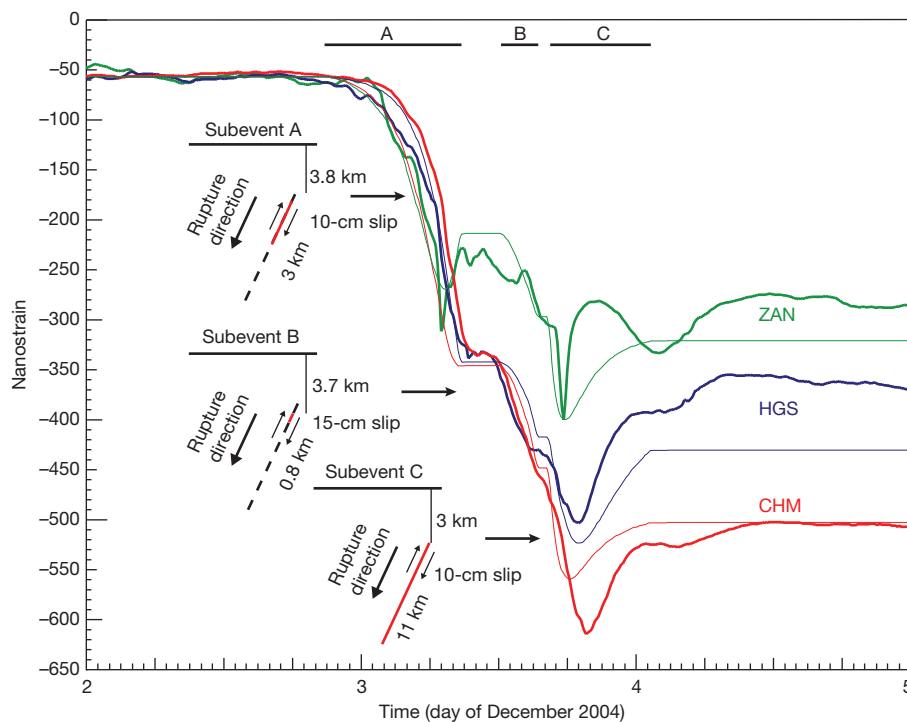


Figure 3 | Data for the same event as in Fig. 2 but after removal of tidal frequencies and atmospheric pressure-induced changes. We divide the event into three stages, indicated by bars A, B and C. Model vertical sections are shown on the left; see Fig. 1 for fault location. Fault lengths are all 3 km.

Reverse slip propagates down dip to different depths for the three stages; deeper slip results in a nodal line sweeping past the sites at different times, producing the reversals in strain. All slips are constant along strike and down dip. This provides a good fit to the data.

for failure as a result of lower atmospheric pressure. Typhoons lower pressure on the land but not on the ocean bottom because pressure decrease on the ocean surface is accompanied by a change in sea level such that pressure at depth remains constant; tide-gauge data confirm the increase in sea level. Thus, lower pressure on land both lowers the normal force across the fault and increases the shear stress on a reverse fault, resulting in a slight increase in the likelihood of rupture.

We identify three episodes of slip in this event (Fig. 3) and model accordingly, looking for solutions with sources on the fault indicated by the seismicity. All changes are very slow, so we use static solutions for deformation due to a buried shear dislocation²⁸ and generate a time series by successive calculations for a model with slow propagation of the rupture⁴. Our model (Fig. 3) is comprised of three subevents all starting at about 3-km depth and propagating to different depths; each event has uniform slip along strike and down dip. All sources extend 3 km in the strike direction (N 190° E) and dip 65° W. We use a horizontal line source with slow propagation down dip. For each of these subevents, all sites initially undergo contraction (negative strain) and the sites surround a surface area of positive strain. As the depth of the line source increases, this positive strain area expands, with ZAN being the first site to experience the sign change. Figure 3 shows both the data and the strain time series that result from the model.

This very simple model gives a remarkably good fit to the data. Importantly, we can find a reasonable and simple model for a slow earthquake that satisfies the data. Episodic slip on an east-dipping fault will satisfy the data equally well; both models lead to the same conclusions. The seismic moment release corresponds to an earthquake with magnitude $M_w \approx 5.4$. This event is the largest and most complex we have observed. Some smaller events are similar but most have a much simpler character with waveform similar to the first subevent in the example analysed. Thus, all event data are consistent with a similar source model. The model for the largest event results in surface displacements that are just below the resolution of continuous Global Positioning System sites in the area, consistent with observation.

We plot the magnitudes of the observed strain events versus the pressure changes of the typhoons in Fig. 4. Our modelling assumes an elastic medium, so the strain amplitudes are proportional to the size

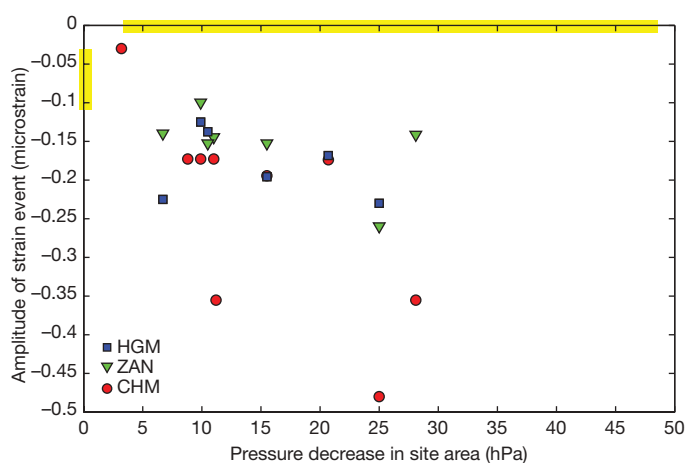


Figure 4 | Amplitudes of the strain signals as a function of pressure drop in the site area as a result of typhoons. The yellow bar along the pressure axis indicates the range of typhoon pressure changes for which no triggered events have been detected; that on the strain axis indicates the size of slow events that occur without typhoon triggering. Scatter in the data set (raw data) prohibit determination of whether event amplitude increases with pressure drop. The CHM data show the largest apparent increase in amplitude; to the extent that that is valid, it is an indication of some variability in the source parameters.

of the slow slip events. On the pressure axis we show, with a yellow bar, the range of pressure drops in the site area for typhoons that do not trigger a slow earthquake; we note that the most severe typhoon did not cause triggering. On the strain axis we show, also with a yellow bar, the range of amplitudes for slow events that occur in the absence of typhoons; all of these are relatively small in comparison with triggered events. There is considerable scatter in the data, with only one site (CHM) apparently showing an increase of strain amplitude with pressure drop. Although more data are necessary before we can investigate this with confidence, it appears that there is some variation of source parameters with the magnitude of the triggering signal.

It has been suggested before that an annual modulation of seismic activity²⁹ (for magnitudes at or below the completeness threshold) is due to the annual change in atmospheric pressure; here we make a definitive connection between fault slip and changes in atmospheric pressure.

We can make a rough estimate of the stress change across the fault interface by assuming an elastic medium. A pressure drop of Δp reduces the vertical stress but the horizontal stress (due to plate convergence) is essentially unaltered because the pressure drop occurs over a large area. This promotes failure both by unclamping ($\Delta p \cos(\theta_{\text{dip}}) \approx 0.8 \text{ kPa}$, reduction in normal force; θ_{dip} , dip angle) and by increasing the reverse shear stress ($\Delta p \sin(\theta_{\text{dip}}) \approx 2 \text{ kPa}$). Stress changes at depth due to seismic waves that cause triggering may be greater by about an order of magnitude¹⁶, so our estimate of the low triggering stress requires that the fault be very close to the failure state; that we do not observe remote seismic triggering suggests that, for a slow earthquake, the trigger duration needs to be longer than that for an ordinary earthquake. At least at the depth of the strainmeters ($\sim 200 \text{ m}$), the strain induced by pressure change is much larger than that of the tides. Our requirement that the fault be critical for triggering is consistent with only a minority of typhoons acting as triggers and slow earthquakes occasionally occurring without a typhoon trigger.

For the slow events not triggered by typhoons, we look for signs of increased levels in background seismic noise as observed in Cascadia⁷ and southwest Japan⁸; typhoon-generated noise precludes any such recognition for the triggered events. We find no such increase in noise level, so these slow events are unaccompanied by a series of long-period earthquakes³⁰.

The total amount of slip (on the fault section common to the subevents) accounts for about four years of plate convergence. The remaining slow events (all much smaller in amplitude), assuming exactly the same source area, require about one to two years of convergence. Although our time window of five years is much too short to enable a meaningful quantitative estimate, it appears that these slow events may be releasing most of the plate convergence on this fault segment.

Our observational program is currently being expanded to have another two small groups of borehole strainmeters, about 90 km to the north and 45 km to the south of the sites involved in this study. From the more northerly area, we have data indicative of another slow earthquake source in that area. The slow events discussed here act to relieve the effects of tectonic compression locally. If there were a number of such sources distributed along this part of the collision zone, they would segment the stressed section such that seismic slip over an extended fault becomes less likely. This may explain why, relative to the high convergence rate (8.2 cm yr^{-1}), there is a paucity of large earthquakes in this area.

Received 8 December 2008; accepted 1 April 2009.

1. Sacks, I. S., Suyehiro, S., Linde, A. T. & Snoko, J. A. Slow earthquakes and stress redistribution. *Nature* **275**, 599–602 (1978).
2. Sacks, I. S., Suyehiro, S., Linde, A. T. & Snoko, J. A. Stress redistribution and slow earthquakes. *Tectonophysics* **81**, 311–318 (1982).
3. Gladwin, M. T., Gwyther, R. L., Hart, R. H. G. & Brechenridge, K. S. Measurements of the strain field associated with episodic creep events on the San Andreas fault at San Juan Bautista, California. *J. Geophys. Res.* **99**, 4559–4564 (1994).

4. Linde, A. T., Gladwin, M. T., Johnston, M. J. S., Gwyther, R. L. & Bilham, R. G. A slow earthquake sequence on the San Andreas fault. *Nature* **383**, 65–68 (1996).
5. Heki, K., Miyazaki, S. & Tsuji, H. Silent fault slip following an interplate thrust earthquake at the Japan Trench. *Nature* **386**, 595–598 (1997).
6. Crescentini, L., Amoruso, A. & Scarpa, R. Constraints on slow earthquake dynamics from a swarm in central Italy. *Science* **286**, 2132–2134 (1999).
7. Dragert, H., Wang, K. & Rogers, G. Geodetic and seismic signatures of episodic tremor and slip in the northern Cascadia subduction zone. *Earth Planets Space* **56**, 1143–1150 (2004).
8. Ito, Y., Obara, K., Shiomi, K., Sekine, S. & Hirose, H. Slow earthquakes coincident with episodic tremors and slow slip events. *Science* **26**, 503–506 (2006).
9. Linde, A. T. & Sacks, I. S. Slow earthquakes and great earthquakes along the Nankai trough. *Earth Planet. Sci. Lett.* **203**, 265–275 (2002).
10. Kato, N. A possible model for large preseismic slip on a deeper extension of a seismic rupture plane. *Earth Planet. Sci. Lett.* **216**, 17–25 (2003).
11. Rydelek, P. A. & Sacks, I. S. Asthenospheric viscosity inferred from correlated land–sea earthquakes in north-east Japan. *Nature* **336**, 234–237 (1988).
12. Pollitz, F. F. & Sacks, I. S. Consequences of stress changes following the 1891 Nobi earthquake, Japan. *Bull. Seismol. Soc. Am.* **85**, 796–807 (1995).
13. Hill, D. P. *et al.* Seismicity remotely triggered by the magnitude 7.3 Landers, California, earthquake. *Science* **250**, 1617–1623 (1993).
14. Brodsky, E., Karakostas, V. & Kanamori, H. A new observation of dynamically triggered regional seismicity: earthquakes in Greece following the August, 1999 Izmit, Turkey earthquake. *Geophys. Res. Lett.* **27**, 2742–2744 (2000).
15. Gombert, J., Bodin, P., Larson, K. & Dragert, H. Earthquake nucleation by transient deformations caused by the $M = 7.9$ Denali, Alaska, earthquake. *Nature* **427**, 621–624 (2004).
16. West, M., Sánchez, J. J. & McNutt, S. R. Periodically triggered seismicity at Mount Wrangell, Alaska, after the Sumatra earthquake. *Science* **308**, 1144–1146 (2005).
17. Miyazawa, M. & Mori, J. Detection of triggered deep low-frequency events from the 2003 Tokachi-oki earthquake. *Geophys. Res. Lett.* **32**, doi:10.1029/2005GL022539 (2005).
18. Suppe, J. Kinematics of arc-continent collision, flipping of subduction, and back-arc spreading near Taiwan. *Mem. Geol. Soc. China* **6**, 21–23 (1984).
19. Tsai, Y. B. Seismotectonics of Taiwan. *Tectonophysics* **125**, 17–38 (1986).
20. Ho, C. S. *An Introduction to the Geology of Taiwan: Explanatory Text of the Geologic Map of Taiwan* 2nd edn (Central Geological Survey, Taiwan, 1986).
21. Chi, W. R., Namson, J. & Suppe, J. Stratigraphic record of plate interactions in the Coastal Range of Eastern Taiwan. *Mem. Geol. Soc. China* **4**, 155–194 (1981).
22. Liu, C. C. & Yu, S. B. Vertical crustal movement in eastern Taiwan and its tectonic implications. *Tectonophysics* **183**, 111–119 (1990).
23. Hsieh, M. L., Liu, P. M. & Hsu, M. Y. Holocene tectonic uplift on the Hua-Tung coast, Eastern Taiwan. *Quaternary Int.* **115–116**, 47–70 (2004).
24. Yu, S. B. & Liu, C. C. Fault creep on the central segment of the Longitudinal Valley fault, Eastern Taiwan. *Proc. Geol. Soc. China* **32**, 209–231 (1989).
25. Lee, J. C., Angelier, J., Chu, H. T., Hu, J. C. & Jeng, F. S. Continuous monitoring of an active fault in a plate suture zone: a creepmeter study of the Chihshang Fault, eastern Taiwan. *Tectonophysics* **333**, 219–240 (2000).
26. Sacks, I. S., Suyehiro, S. & Evertson, D. W. Sacks–Evertson strainmeter, its installation in Japan and some preliminary results concerning strain steps. *Proc. Jpn Acad.* **47**, 707–712 (1971).
27. Kim, K. H. *et al.* Three-dimensional V_P and V_S structural models associated with the active subduction and collision tectonics in the Taiwan region. *Geophys. J. Int.* **162**, 204–220 (2005).
28. Okada, Y. Internal deformation due to shear and tensile faults in a half-space. *Bull. Seismol. Soc. Am.* **82**, 1018–1040 (1992).
29. Gao, S., Silver, P. G., Linde, A. T. & Sacks, I. S. Annual modulation of triggered seismicity following the 1992 Landers earthquake in California. *Nature* **406**, 500–504 (2000).
30. Shelly, D. R., Beroza, G. C., Ide, S. & Nakamura, S. Low-frequency earthquakes in Shikoku, Japan, and their relationship to episodic tremor and slip. *Nature* **442**, 188–191 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Lee for critically important work in initiating this project. M. Lee's expertise was valuable in site selection. J.-C. Rau and T.-S. Lee gave access to their property for borehole sites. H. M. Lee managed the archiving of data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.L. (liucc@sinica.edu.tw).

Two types of dopamine neuron distinctly convey positive and negative motivational signals

Masayuki Matsumoto¹ & Okihide Hikosaka¹

Midbrain dopamine neurons are activated by reward or sensory stimuli predicting reward^{1–4}. These excitatory responses increase as the reward value increases⁵. This response property has led to a hypothesis that dopamine neurons encode value-related signals and are inhibited by aversive events. Here we show that this is true only for a subset of dopamine neurons. We recorded the activity of dopamine neurons in monkeys (*Macaca mulatta*) during a Pavlovian procedure with appetitive and aversive outcomes (liquid rewards and airpuffs directed at the face, respectively). We found that some dopamine neurons were excited by reward-predicting stimuli and inhibited by airpuff-predicting stimuli, as the value hypothesis predicts. However, a greater number of dopamine neurons were excited by both of these stimuli, inconsistent with the hypothesis. Some dopamine neurons were also excited by both rewards and airpuffs themselves, especially when they were unpredictable. Neurons excited by the airpuff-predicting stimuli were located more dorsolaterally in the substantia nigra pars compacta, whereas neurons inhibited by the stimuli were located more ventromedially, some in the ventral tegmental area. A similar anatomical difference was observed for their responses to actual airpuffs. These findings suggest that different groups of dopamine neurons convey motivational signals in distinct manners.

If midbrain dopamine neurons actually encode value-related signals, their activity should be inhibited by aversive stimuli because aversive stimuli have negative motivational values. However, the results are inconsistent, some studies showing inhibitions⁶ and others showing both inhibitions and excitations^{7–11} by aversive stimuli. Few of these studies examined the effects of rewards on the same dopamine neurons^{12,13}, partly because the animals were anaesthetized.

To test whether dopamine neurons encode motivational values, we conditioned two monkeys using a Pavlovian procedure with two distinct contexts (Fig. 1): one in which a liquid reward was expected (appetitive block; Fig. 1a) and one in which an aversive airpuff was anticipated (aversive block; Fig. 1b). In each block, three conditioned stimuli were associated with the unconditioned stimulus (reward or airpuff) with 100%, 50% and 0% probability, respectively. These three conditioned stimuli were considered to convey three different levels of motivational value. In the appetitive block, anticipatory licking increased as the probability of reward increased (Fig. 1c). In the aversive block, anticipatory blinking increased as the probability of airpuff increased (Fig. 1d).

While the monkeys were conditioned using the Pavlovian procedure, we recorded single-unit activity from 103 putative dopamine neurons (68 in monkey N and 35 in monkey D) in and around the substantia nigra pars compacta (SNc) and ventral tegmental area (VTA). Their electrophysiological properties were distinctly different from other neurons in the SNc and VTA (Supplementary Fig. 1), and we henceforth call them dopamine neurons.

Most previous studies on midbrain dopamine neurons have characterized dopamine neurons as a functionally homogeneous population¹.

We found that this is not true. In Fig. 2a, e, we show the activity of two dopamine neurons, separately for different conditioned stimuli. Their

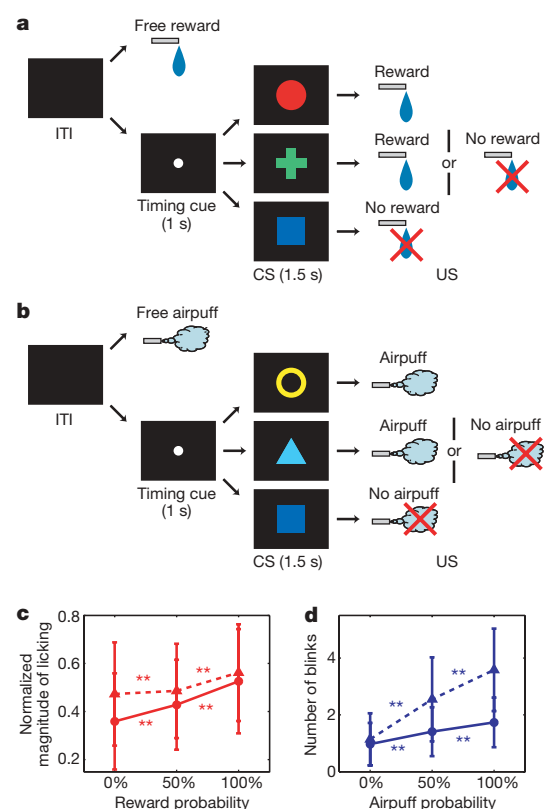


Figure 1 | Pavlovian procedure. **a**, Appetitive block. Three conditioned stimuli were associated with apple juice with 100%, 50% and 0% probability, respectively. **b**, Aversive block. Three conditioned stimuli were associated with an aversive airpuff with 100%, 50% and 0% probability, respectively. In both blocks, each trial started after the presentation of a timing cue (central small spot) on the screen. After 1 s, the timing cue disappeared and one of the three conditioned stimuli was presented. After 1.5 s, the conditioned stimulus disappeared and the unconditioned stimulus (reward or airpuff) was delivered. In addition to the cued trials, uncued trials were included in which a reward alone (free reward) was delivered during the appetitive block and an airpuff alone (free airpuff) was delivered during the aversive block. **c**, Average normalized magnitude of anticipatory licking during the presentation of the reward-predicting conditioned stimuli for monkey D (solid line) and monkey N (dashed line). **d**, Average number of anticipatory blinks during the presentation of the airpuff-predicting conditioned stimuli for monkey D (solid line) and monkey N (dashed line). Double asterisks indicate a significant difference between two data points ($P < 0.01$, Wilcoxon rank-sum test). Error bars, s.d. ITI, inter-trial interval; CS, conditioned stimulus; US, unconditioned stimulus.

¹Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, Maryland 20892-4435, USA.

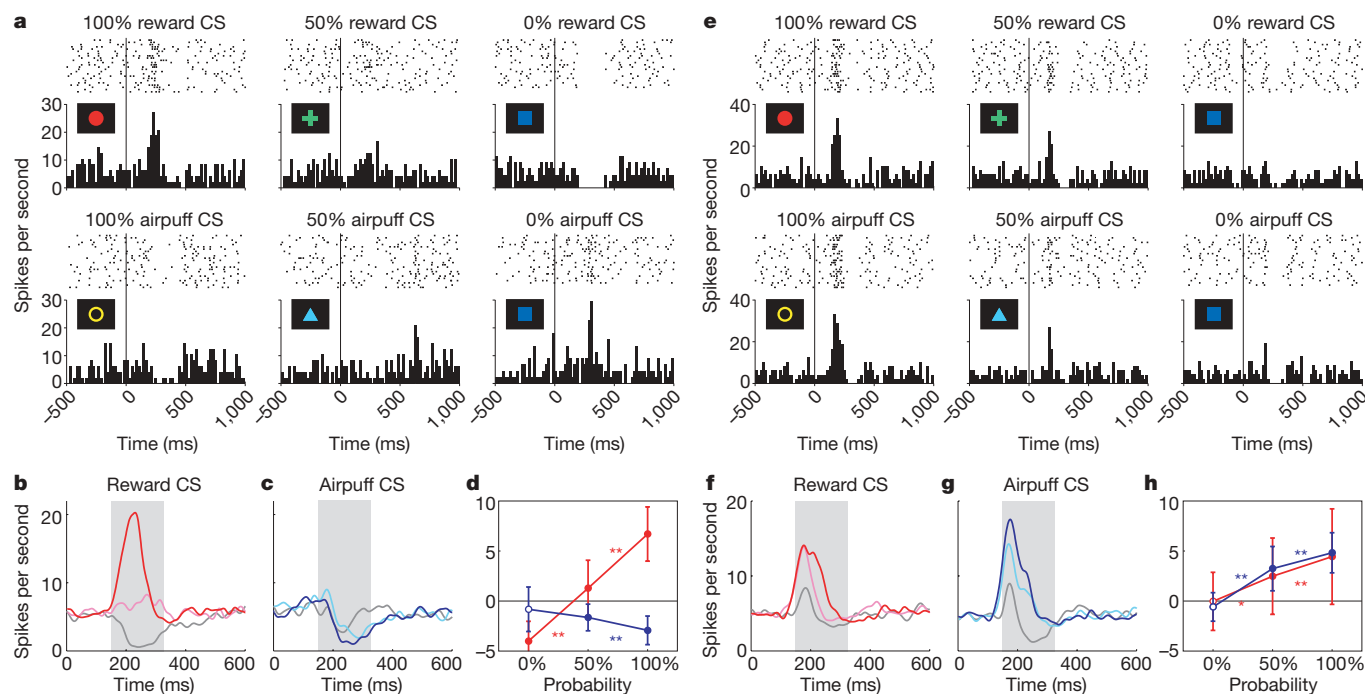


Figure 2 | Responses of dopamine neurons to conditioned stimuli.

a, e, Activity of two example neurons in the appetitive block (top row) and aversive block (bottom row), which were classified as ACS-inhibited type (**a**) and ACS-excited type (**e**). Histograms (20-ms bins) and rasters are aligned at the start of the conditioned stimulus and are shown for 100% reward CS, 50% reward CS, 0% reward CS, 100% airpuff CS, 50% airpuff CS and 0% airpuff CS. **b, c,** Averaged activity of 24 ACS-inhibited neurons. **f, g,** Averaged activity of 38 ACS-excited neurons. Spike density functions are shown for 100% reward CS (red), 50% reward CS (pink) and 0% reward CS (grey) in the appetitive block (**b, f**), and for 100% airpuff CS (dark blue), 50%

activities were similar in the appetitive block (top row). Both of them were excited by 100% reward conditioned stimulus (the conditioned stimulus associated with reward with 100% probability). This excitation decreased in response to 50% reward conditioned stimulus, and changed to an inhibition in response to 0% reward conditioned stimulus. However, the dopamine neurons showed completely different responses in the aversive block (bottom row). In response to 100% airpuff conditioned stimulus, the neuron shown in Fig. 2a was inhibited whereas the neuron shown in Fig. 2e was excited. Furthermore, as the probability of airpuff decreased, their response magnitudes were graded in opposite directions.

To characterize the responses to conditioned stimuli, we classified the 103 neurons into three groups based on the response to 100% airpuff conditioned stimulus (Supplementary Table 1). Neurons showing a significant inhibition and excitation were classified as airpuff conditioned stimulus (ACS)-inhibited type ($n = 24$) and ACS-excited type ($n = 38$), respectively ($P < 0.05$, Wilcoxon signed-rank test). Neurons showing no significant response were classified as ACS-non-responsive type ($n = 41$, $P > 0.05$, Wilcoxon signed-rank test). The responses of individual neurons to conditioned stimuli are shown in Supplementary Fig. 2 and Supplementary Table 2. In the following, we will focus on the ACS-inhibited and ACS-excited neurons (see Supplementary Fig. 3 for ACS-non-responsive neurons; see also Supplementary Note A and Supplementary Table 3 for the electrophysiological properties of each type).

The averaged activity of the ACS-inhibited neurons was modulated by the reward probability (Fig. 2b) and airpuff probability (Fig. 2c) in opposite directions. The excitatory response to the reward-predicting conditioned stimuli decreased and became an inhibition as the reward probability decreased (Fig. 2b, red line in Fig. 2d). By contrast, the inhibitory response to the airpuff-predicting conditioned stimuli decreased as the airpuff probability decreased

(Fig. 2c, blue line in Fig. 2d). The same trend was found in individual ACS-inhibited neurons (Supplementary Note B and Supplementary Fig. 4a). These results suggest that the ACS-inhibited neurons encode motivational value on a single scale, and are most strongly excited in response to the most positive stimulus (100% reward conditioned stimulus) and most strongly inhibited in response to the most negative stimulus (100% airpuff conditioned stimulus).

The averaged activity of the ACS-excited neurons was also modulated by the reward probability (Fig. 2f) and airpuff probability (Fig. 2g), but in the same direction. The excitatory response decreased as the outcome probability decreased for both reward-predicting and airpuff-predicting conditioned stimuli (Fig. 2h; see also Supplementary Note B and Supplementary Fig. 4b for individual neurons). These results suggest that the ACS-excited neurons do not encode motivational value.

Previous studies have repeatedly shown that dopamine neurons are excited by reward when it is unexpected¹. However, it is still debatable whether they are excited or inhibited by aversive stimuli and, if so, in what context. Figure 3a shows the responses to reward and airpuff of the same neuron shown in Fig. 2a. This neuron was strongly excited when reward was presented without preceding conditioned stimulus (free reward) and inhibited when airpuff was presented without preceding conditioned stimulus (free airpuff), consistent with value coding. By contrast, the neuron shown in Fig. 3e was excited by both free reward and free airpuff.

We then reclassified the 103 neurons into three groups on the basis of the response to free airpuff (Supplementary Table 1). Neurons showing significant inhibition and excitation were classified as airpuff unconditioned stimulus (AUS)-inhibited type ($n = 47$) and AUS-excited type ($n = 11$), respectively ($P < 0.05$, Wilcoxon signed-rank test). Neurons showing no significant response were classified as AUS-non-responsive type ($n = 45$, $P > 0.05$, Wilcoxon signed-rank test). The

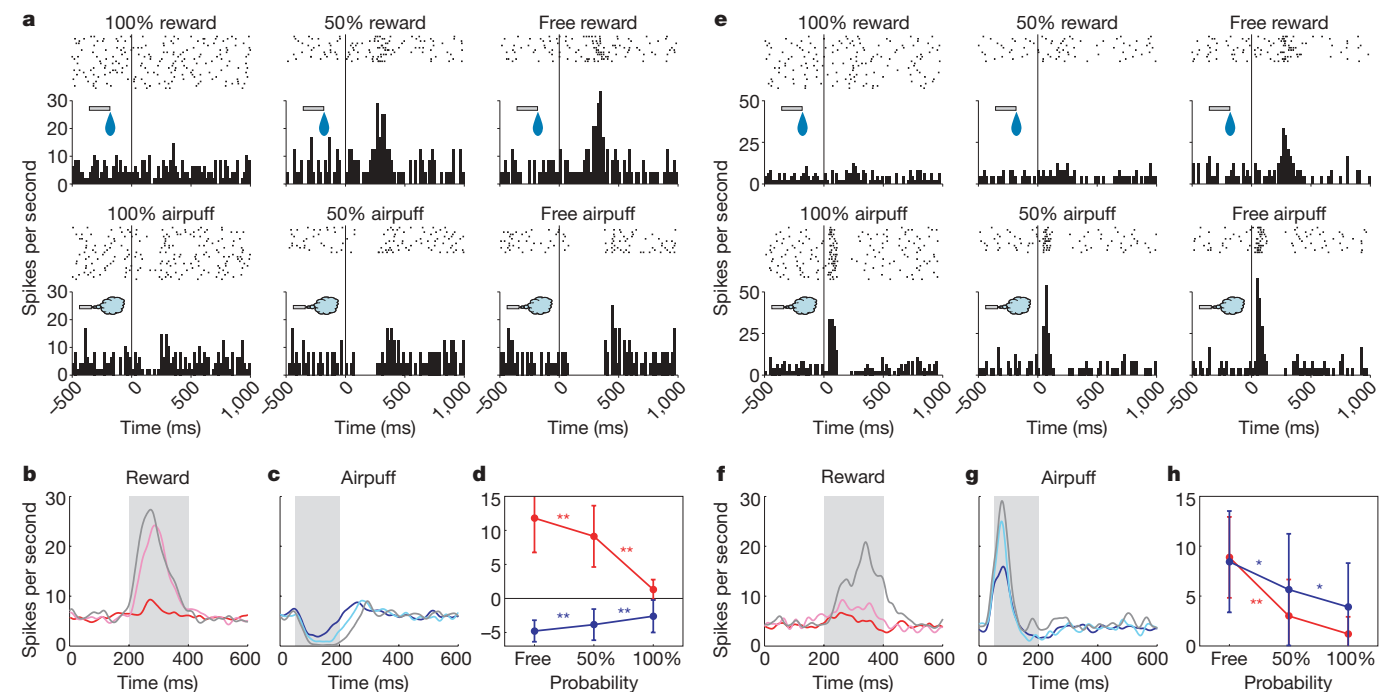


Figure 3 | Responses of dopamine neurons to unconditioned stimuli.

a, e, Activity of two example neurons in the appetitive block (top row) and aversive block (bottom row), which were classified as AUS-inhibited type (**a**) and AUS-excited type (**e**). Histograms and rasters are aligned at the start of the unconditioned stimulus and are shown for 100% reward, 50% reward, free reward, 100% airpuff, 50% airpuff and free airpuff. **b, c**, Averaged activity of 47 AUS-inhibited neurons. **f, g**, Averaged activity of 11 AUS-excited neurons. Spike density functions are shown for 100% reward (red),

50% reward (pink) and free reward (grey) in the appetitive block (**b, f**), and for 100% airpuff (dark blue), 50% airpuff (light blue) and free airpuff (grey) in the aversive block (**c, g**). Grey areas indicate the period that was used to analyse responses to unconditioned stimuli. **d, h**, The magnitudes of the responses of the AUS-inhibited neurons (**d**) and AUS-excited neurons (**h**) to reward (red) and airpuff (blue). Significance measures and error bars are the same as Fig. 2d, h.

responses of individual neurons to unconditioned stimuli are shown in Supplementary Fig. 5 and Supplementary Table 2. We note that this classification differs from that based on the response to 100% airpuff conditioned stimulus. In the following, we will focus on the AUS-inhibited and AUS-excited neurons (see Supplementary Figs 6 and 7 for AUS-non-responsive neurons, see also Supplementary Note C and Supplementary Table 4 for the electrophysiological properties of each type).

The averaged responses to the reward and airpuff are shown for the AUS-inhibited neurons in Fig. 3b, c and for the AUS-excited neurons in Fig. 3f, g. In both types, the excitatory response to reward disappeared when the reward was completely predictable by following 100% reward conditioned stimulus, and decreased when the reward was partly predictable by following 50% reward conditioned stimulus (Fig. 3b, f). This is consistent with the reward-prediction-error hypothesis that the activity of dopamine neurons represents the difference between the expected and actual values of reward^{14,15}.

The AUS-inhibited neurons appeared to encode prediction error even for aversive outcomes, albeit partly, because these neurons were inhibited by an unexpected aversive airpuff (free airpuff; Fig. 3c) and this inhibitory response decreased monotonically as the airpuff became more predictable (Fig. 3d; see Supplementary Note D and Supplementary Fig. 8a for individual neurons). We note that the excitatory response of the AUS-excited neurons to the airpuff also decreased as the airpuff became more predictable (Fig. 3g, h; see Supplementary Note D and Supplementary Fig. 8b for individual neurons).

The prediction-error hypothesis predicts that when an outcome is unexpectedly omitted, neurons should respond in the direction opposite to that in which they respond when the same outcome is unexpectedly delivered^{14,15}. We found that AUS-inhibited neurons tended to show this kind of response to both reward omission and

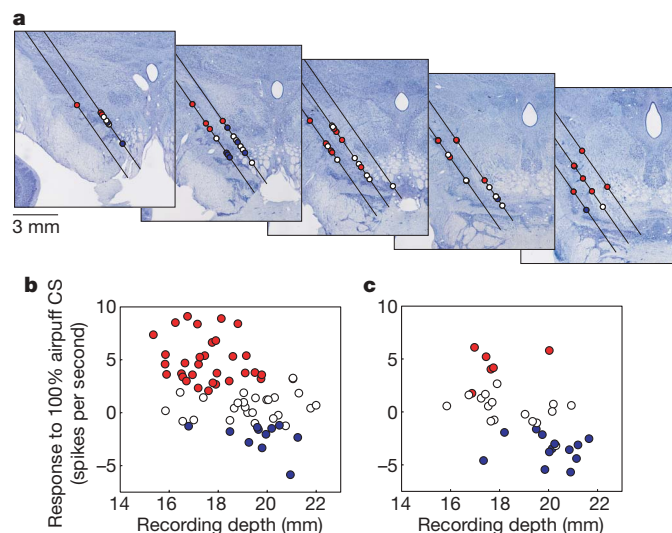


Figure 4 | Locations of dopamine neurons in relation to their responses to airpuff-predicting conditioned stimulus. **a**, Recording sites of 68 dopamine neurons in monkey N are plotted on five coronal sections shown rostrocaudally from left to right (interval, 1 mm). Red circles indicate neurons showing significant excitations to 100% airpuff CS (that is, ACS-excited neurons). Blue circles indicate neurons showing significant inhibitions to 100% airpuff CS (that is, ACS-inhibited neurons). White circles, no significance (that is, ACS-non-responsive neurons). Black lines indicate electrode penetration tracks, which were tilted laterally by 35°. **b, c**, Relation between recording depth and the response to 100% airpuff CS for monkey N (**b**) and monkey D (**c**). Red, blue and white circles indicate ACS-excited, ACS-inhibited and ACS-non-responsive neurons, respectively. The recording depth was measured from a reference depth set by a manipulator to advance the recording electrode.

airpuff omission, whereas AUS-excited neurons showed no response to the omission of the outcome (Supplementary Note E and Supplementary Fig. 9).

The current consensus, that dopamine neurons carry reward-related information, is thought to hold for all dopamine neurons located in the midbrain, including both the SNc and the VTA¹. Because we have found types of dopamine neuron that differ with regard to their responses to aversive events, we now ask whether they were located in different regions in the midbrain. Figure 4a shows the recording sites of the 68 dopamine neurons in monkey N in relation to the response to 100% airpuff conditioned stimulus. Neurons showing a significant excitation (that is, ACS-excited neurons; red circles) tended to be located in the more dorsolateral part, and neurons showing a significant inhibition (that is, ACS-inhibited neurons; blue circles) tended to be located in the more ventromedial part. To test this trend statistically, we examined the relation between the recording depth and the response to 100% airpuff conditioned stimulus for monkey N (Fig. 4b) and monkey D (Fig. 4c). As shown by the scatter plots, a significant negative correlation was found for both monkeys (monkey N: correlation coefficient $r = -0.50$, $P < 0.01$; monkey D: $r = -0.57$, $P < 0.01$). This negative correlation confirmed the dorsolateral–ventromedial differentiation of the excitatory and inhibitory responses evoked in dopamine neurons by the airpuff-predicting conditioned stimulus. Similar location differences were found in relation to response to airpuff itself (Supplementary Note F and Supplementary Fig. 10).

It has generally been assumed that midbrain dopamine neurons form a unified functional group, all representing reward-related signals in a similar manner¹. Our results are roughly consistent with this idea as far as the reward-related signals are concerned. However, clear heterogeneity was revealed when we examined their responses to aversive events. We found two types of dopamine neuron, one inhibited and the other excited by airpuff or its predictor. This suggests that the unified concept of dopamine neurons needs to be changed (see Supplementary Note G for the relationship between our findings and previous studies).

We propose that there are at least two functional groups of dopamine neurons. Dopamine neurons in the first group (airpuff-inhibited type, that is, ACS- and AUS-inhibited types) would represent motivational value. Their responses co-varied with prediction errors associated with both reward and airpuff, and therefore would be useful in learning to approach rewards and avoid aversive stimuli. The function of the second group (airpuff-excited type, that is, ACS- and AUS-excited types) is not immediately clear, but we found that their response to the conditioned stimulus was correlated with the latency of the monkey's orienting response (gaze shift) to the conditioned stimulus and that this correlation appeared only after the conditioned stimulus was paired with reward or airpuff (Supplementary Note H and Supplementary Fig. 11). These results raise the possibility that the responses of the airpuff-excited dopamine neurons to a conditioned stimulus reflect the motivational salience of the conditioned stimulus. However, this interpretation may not be valid for the responses of these neurons to unconditioned stimulus or its omission.

We note that the two types of dopamine neuron were distributed differently, the airpuff-excited type in the dorsolateral region in the SNc and the airpuff-inhibited type in the ventromedial region in the SNc as well as the VTA (see Supplementary Note I for details). In monkeys¹⁶ and rats¹⁷, dopamine neurons in the dorsolateral SNc project mainly to the dorsal striatum, whereas those in the ventromedial SNc and VTA project mainly to the ventral striatum. The airpuff-inhibited dopamine neurons in the ventromedial region in the SNc and VTA may thus transmit value-related information to the ventral striatum, which is thought to process reward values^{18–20}. On the other hand, the airpuff-excited dopamine neurons in the dorsolateral region in the SNc respond to motivationally salient stimuli, whether they are appetitive or aversive, and send the signal to the dorsal striatum, which is related to orienting behaviour^{21–23}. This may

be part of the mechanism by which orienting behaviour such as saccadic eye movement is induced by motivationally salient stimuli²⁴.

The two types of dopamine neuron may receive inputs from different sources. The airpuff-excited dopamine neurons may receive inputs from areas such as the basal forebrain, in which neurons also show excitatory responses to both appetitive and aversive events^{25,26} (see Supplementary Note J for further discussion). The airpuff-inhibited dopamine neurons may receive inputs, at least partly, from the lateral habenula. Using the same Pavlovian procedure, we have shown that lateral habenula neurons are excited by the airpuff-predicting conditioned stimulus and inhibited by the reward-predicting conditioned stimulus, indicating that they encode motivational value similarly to the airpuff-inhibited dopamine neurons, but in the opposite manner²⁷. The value signals in the lateral habenula would then be transmitted to the dopamine neurons by inhibiting them²⁸, and this effect was stronger on dopamine neurons located in the ventromedial SNc or the VTA, where the airpuff-inhibited type dominates (Supplementary Note K and Supplementary Fig. 12).

So far, we have classified dopamine neurons into two types. However, the real picture is more complex. First, the difference between the two types was not distinct; there was another group of dopamine neurons that did not belong to either type (that is, the type non-responsive to airpuff or its predictor). Second, the classification was different for conditioned and unconditioned stimuli (Supplementary Note L, Supplementary Table 1 and Supplementary Fig. 13b). More neurons were excited by the airpuff-predicting conditioned stimulus, whereas more neurons were inhibited by the airpuff itself. This might indicate flexible operation of the dopamine system. If a salient stimulus (that is, a conditioned stimulus) is presented, it would be beneficial to orient attention to the stimulus and judge whether it predicts a rewarding event or an aversive event. This is the time when a majority of dopamine neurons are excited, thus promoting the orienting behaviour. If an aversive event occurs (that is, unconditioned stimulus), it would be crucial to learn to avoid the action that led to the aversive event. This is the time when a majority of dopamine neurons are inhibited, thus promoting avoidance learning.

METHODS SUMMARY

Two adult rhesus monkeys (*Macaca mulatta*) were used for the experiments. All procedures for animal care and experimentation were approved by the Animal Care and Use Committee of the National Eye Institute and complied with the Public Health Service Policy on the humane care and use of laboratory animals.

A plastic head holder and plastic recording chamber were fixed to the skull under general anaesthesia and sterile surgical conditions. The recording chamber was placed over the frontoparietal cortex, tilted laterally by 35°, and aimed at the SNc and VTA. Two search coils were surgically placed under the conjunctiva of the eyes. The head holder, the recording chamber and the eye coil connectors were all embedded in dental acrylic that covered the top of the skull, and were connected to the skull using acrylic screws.

We conditioned two monkeys using a Pavlovian procedure with an appetitive unconditioned stimulus (liquid reward) and an aversive unconditioned stimulus (airpuff). During the Pavlovian procedure, we recorded the activity of dopamine neurons in and around the SNc and VTA. We estimated the position of the SNc and VTA by magnetic resonance imaging and identified dopamine neurons by their electrophysiological properties. After the end of recording sessions in one monkey, we confirmed the recording sites histologically. We analysed anticipatory licking, anticipatory blinking and neuronal responses during the Pavlovian procedure. We focused on three kinds of neuronal responses: (1) responses elicited by conditioned-stimulus presentation, (2) responses elicited by unconditioned-stimulus delivery and (3) responses elicited by unconditioned-stimulus omission. Details of the Pavlovian procedure, identification of dopamine neurons, analysis methods, and histological procedure can be found in the full Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 January; accepted 27 March 2009.

Published online 17 May 2009.

- Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).

2. Satoh, T., Nakai, S., Sato, T. & Kimura, M. Correlated coding of motivation and outcome of decision by dopamine neurons. *J. Neurosci.* **23**, 9913–9923 (2003).
3. Takikawa, Y., Kawagoe, R. & Hikosaka, O. A possible role of midbrain dopamine neurons in short- and long-term adaptation of saccades to position-reward mapping. *J. Neurophysiol.* **92**, 2520–2529 (2004).
4. Morris, G., Arkadir, D., Nevet, A., Vaadia, E. & Bergman, H. Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron* **43**, 133–143 (2004).
5. Tobler, P. N., Fiorillo, C. D. & Schultz, W. Adaptive coding of reward value by dopamine neurons. *Science* **307**, 1642–1645 (2005).
6. Ungless, M. A., Magill, P. J. & Bolam, J. P. Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science* **303**, 2040–2042 (2004).
7. Chiodo, L. A., Antelman, S. M., Caggiula, A. R. & Lineberry, C. G. Sensory stimuli alter the discharge rate of dopamine (DA) neurons: evidence for two functional types of DA cells in the substantia nigra. *Brain Res.* **189**, 544–549 (1980).
8. Coizet, V., Dommett, E. J., Redgrave, P. & Overton, P. G. Nociceptive responses of midbrain dopaminergic neurones are modulated by the superior colliculus in the rat. *Neuroscience* **139**, 1479–1493 (2006).
9. Schultz, W. & Romo, R. Responses of nigrostriatal dopamine neurons to high-intensity somatosensory stimulation in the anesthetized monkey. *J. Neurophysiol.* **57**, 201–217 (1987).
10. Mantz, J., Thierry, A. M. & Glowinski, J. Effect of noxious tail pinch on the discharge rate of mesocortical and mesolimbic dopamine neurons: selective activation of the mesocortical system. *Brain Res.* **476**, 377–381 (1989).
11. Guarraci, F. A. & Kapp, B. S. An electrophysiological characterization of ventral tegmental area dopaminergic neurons during differential Pavlovian fear conditioning in the awake rabbit. *Behav. Brain Res.* **99**, 169–179 (1999).
12. Mirenowicz, J. & Schultz, W. Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature* **379**, 449–451 (1996).
13. Joshua, M., Adler, A., Mitelman, R., Vaadia, E. & Bergman, H. Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials. *J. Neurosci.* **28**, 11673–11684 (2008).
14. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
15. Montague, P. R., Dayan, P. & Sejnowski, T. J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**, 1936–1947 (1996).
16. Lynd-Balta, E. & Haber, S. N. The organization of midbrain projections to the striatum in the primate: sensorimotor-related striatum versus ventral striatum. *Neuroscience* **59**, 625–640 (1994).
17. Ikemoto, S. Dopamine reward circuitry: two projection systems from the ventral midbrain to the nucleus accumbens–olfactory tubercle complex. *Brain Res. Rev.* **56**, 27–78 (2007).
18. Knutson, B., Adams, C. M., Fong, G. W. & Hommer, D. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J. Neurosci.* **21**, RC159 (2001).
19. Cromwell, H. C. & Schultz, W. Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. *J. Neurophysiol.* **89**, 2823–2838 (2003).
20. Schultz, W., Apicella, P., Scarnati, E. & Ljungberg, T. Neuronal activity in monkey ventral striatum related to the expectation of reward. *J. Neurosci.* **12**, 4595–4610 (1992).
21. Kitama, T., Ohno, T., Tanaka, M., Tsubokawa, H. & Yoshida, K. Stimulation of the caudate nucleus induces contraversive saccadic eye movements as well as head turning in the cat. *Neurosci. Res.* **12**, 287–292 (1991).
22. Hikosaka, O., Takikawa, Y. & Kawagoe, R. Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiol. Rev.* **80**, 953–978 (2000).
23. Carli, M., Evenden, J. L. & Robbins, T. W. Depletion of unilateral striatal dopamine impairs initiation of contralateral actions and not sensory attention. *Nature* **313**, 679–682 (1985).
24. Holland, P. C. & Gallagher, M. Amygdala circuitry in attentional and representational processes. *Trends Cogn. Sci.* **3**, 65–73 (1999).
25. Lin, S. C. & Nicolelis, M. A. Neuronal ensemble bursting in the basal forebrain encodes salience irrespective of valence. *Neuron* **59**, 138–149 (2008).
26. Richardson, R. T. & DeLong, M. R. Electrophysiological studies of the functions of the nucleus basalis in primates. *Adv. Exp. Med. Biol.* **295**, 233–252 (1991).
27. Matsumoto, M. & Hikosaka, O. Representation of negative motivational value in the primate lateral habenula. *Nature Neurosci.* **12**, 77–84 (2009).
28. Matsumoto, M. & Hikosaka, O. Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* **447**, 1111–1115 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Hong, E. Bromberg-Martin, M. Yasuda, S. Yamamoto and Y. Tachibana for discussion, M. K. Smith for his histological expertise, J. W. McClurkin, A. M. Nichols, T. W. Ruffner, A. V. Hays and L. P. Jensen for technical assistance, and G. Tansey, D. Parker and B. Nagy for animal care. This research was supported by the Intramural Research Program at the National Institutes of Health, National Eye Institute.

Author Contributions M.M. designed the Pavlovian procedure, performed the experiments and analysed the data. O.H. supported all of these processes. M.M. and O.H. discussed the results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.M. (matsumotom@nei.nih.gov).

METHODS

Pavlovian procedure. Our Pavlovian procedure consisted of two blocks of trials, an appetitive block (Fig. 1a) and an aversive block (Fig. 1b). In the appetitive block, three conditioned stimuli (red circle, green cross and blue square for monkey N; yellow ring, cyan triangle and blue square for monkey D) were associated with a liquid reward (apple juice) as an unconditioned stimulus with 100%, 50% and 0% probability, respectively. In the aversive block, three conditioned stimuli (yellow ring, cyan triangle and blue square for monkey N; red circle, green cross and blue square for monkey D) were associated with an airpuff directed at the monkey's face as an unconditioned stimulus with 100%, 50% and 0% probability, respectively. The liquid reward was delivered through a spout that was positioned in front of the monkey's mouth. The airpuff (20–30 p.s.i.) was delivered through a narrow tube placed 6–7 cm from the face.

Each trial started after the presentation of a timing cue for both blocks. The monkeys were not required to look at the timing cue. After 1 s, the timing cue disappeared and one of the three conditioned stimuli was presented pseudo-randomly. After 1.5 s, the conditioned stimulus disappeared and the unconditioned stimulus was delivered. In addition to the cued trials, uncued trials were included in which a reward alone (free reward) was delivered during the appetitive block and an airpuff alone (free airpuff) was delivered during the aversive block. All trials were presented with a random inter-trial interval that averaged 5 s (3–7 s) for monkey N and 4.5 s (3–6 s) for monkey D. One block consisted of 42 trials with fixed proportions of trial types (100%, 12 trials; 50%, 12 trials; 0%, 12 trials; uncued, 6 trials). For 50% trials, the conditioned stimulus was followed by the unconditioned stimulus in six trials and was not followed by the unconditioned stimulus in the other six trials. The block changed without any external cue. For each neuron, we collected data by repeating the appetitive and aversive blocks twice or more.

We monitored licking and blinking of the monkeys. To monitor licking, we attached a strain gauge to the reward spout and measured strains on the spout resulting from licking. To monitor blinking, a magnetic-search-coil technique was used. A small Teflon-coated stainless-steel wire (<5 mm in diameter, five or six turns) was taped to an eyelid. Eye closure was identified by the vertical component of the eyelid-coil signal.

Identification of dopamine neurons. We searched for dopamine neurons in and around the SNc and VTA. Dopamine neurons were identified by their irregular firing, tonic baseline activity around five spikes per second, broad spike potential and phasic excitation to free reward.

Data analysis. We analysed anticipatory licking, anticipatory blinking and neuronal activity during the Pavlovian procedure.

To evaluate the frequency and strength of anticipatory licking, the strain-gauge signal was used. We first calculated the velocity of the signal change under licking. Then we integrated the absolute velocity during conditioned-stimulus presentation for each trial. This integrated velocity becomes larger if the monkeys more frequently and strongly lick the spout. We defined this value as the magnitude of anticipatory licking in the trial. The magnitude was normalized according to the following formula: normalized magnitude equals $(X - \text{Min}) / (\text{Max} - \text{Min})$. Here X is the magnitude of anticipatory licking in the trial, Max is the maximum magnitude in the recording session and Min is the minimum magnitude in the recording session.

To count the number of anticipatory blinks during conditioned-stimulus presentation, the vertical component of the eyelid signal was used. We first calculated the downward velocity of eyelid movement. We set a threshold and counted how many times the velocity crossed the threshold during conditioned-stimulus

presentation for each trial. This count was defined as the number of anticipatory blinks in the trial.

In analyses of neuronal activity, responses to each conditioned stimulus were defined as the discharge rate during the interval 150 to 325 ms after conditioned stimulus onset minus the background discharge rate during the 250 ms before conditioned stimulus onset. Response to reward was defined as the discharge rate during the interval 200 to 400 ms after reward onset minus the background discharge rate during the 250 ms before reward onset. Response to airpuff was defined as the discharge rate during the interval 50 to 200 ms after airpuff onset minus the background discharge rate during the 250 ms before airpuff onset. Response to reward omission was defined as the discharge rate during the interval 200 to 500 ms after the conditioned stimulus ended minus the background discharge rate during the 250 ms before the conditioned stimulus ended. Response to airpuff omission was defined as the discharge rate during the interval 150 to 350 ms after the conditioned stimulus ended minus the background discharge rate during the 250 ms before the conditioned stimulus ended. These time windows were determined on the basis of the averaged activity of dopamine neurons. Specifically, we set the time windows such that they include major parts of the excitatory and inhibitory responses.

Because the 0% reward conditioned stimulus and 0% airpuff conditioned stimulus were physically identical, they could only be distinguished by the block context (appetitive block or aversive block). Therefore, to analyse responses to 0% reward conditioned stimulus and 0% airpuff conditioned stimulus, we excluded all trials with the 0% reward conditioned stimulus or the 0% airpuff conditioned stimulus that were presented before the block context could be known, that is, before the block's first presentation of 100% conditioned stimulus, 50% conditioned stimulus or free outcome.

We characterized the electrophysiological properties of recorded neurons by (1) baseline firing rate, (2) irregularity of firing pattern and (3) spike waveform. Baseline firing rate is the mean firing rate during the 250 ms before the onset of the timing cue. To quantify irregularity of firing pattern, we used an irregularity metric introduced in ref. 29 and called 'IR'. First, interspike interval (ISI) was computed as follows: if spike $i - 1$, spike i and spike $i + 1$ occurred in this order, the interval between spike $i - 1$ and spike i corresponds to ISI_i , and the interval between spike i and spike $i + 1$ corresponds to ISI_{i+1} . Second, the difference between adjacent ISIs was computed as $|\log(\text{ISI}_i / \text{ISI}_{i+1})|$. This value was then assigned to the time spike i occurred. Thus, small IR values indicate regular firing and large IR values indicate irregular firing. We then computed a median of all IR values during the inter-trial interval (during the 1,000 ms before timing-cue onset). To quantify spike waveform, we measured the spike duration of 67 dopamine neurons (whose spike waveforms were successfully recorded). The typical spike consisted of the following waves: first, sharp negative; second, sharp positive; third, slow negative; fourth, slow positive. We measured the spike duration from the peak of the first wave (sharp negative) to the peak of the third wave (slow negative).

Histology. After the end of the recording session in monkey N, we selected representative locations for electrode penetration. When typical dopamine activity was recorded, we made electrolytic microlesions at the recording sites (12 μA and 30 s). Then monkey N was deeply anaesthetized using pentobarbital sodium, and perfused with 10% formaldehyde. The brain was blocked and equilibrated with 10% sucrose. Frozen sections were cut every 50 μm in the coronal plane. The sections were stained with cresyl violet.

29. Davies, R. M., Gerstein, G. L. & Baker, S. N. Measurement of time-dependent changes in the irregularity of neural spiking. *J. Neurophysiol.* **96**, 906–918 (2006).

LETTERS

Specificity of sensory–motor connections encoded by *Sema3e*–*Plxnd1* recognition

Eline Pecho-Vrieseling¹, Markus Sigrist¹, Yutaka Yoshida^{2,3}, Thomas M. Jessell² & Silvia Arber¹

Spinal reflexes are mediated by synaptic connections between sensory afferents and motor neurons^{1–3}. The organization of these circuits shows several levels of specificity. Only certain classes of proprioceptive sensory neurons make direct, monosynaptic connections with motor neurons⁴. Those that do are bound by rules of motor pool specificity: they form strong connections with motor neurons supplying the same muscle, but avoid motor pools supplying antagonistic muscles^{1,5–7}. This pattern of connectivity is initially accurate and is maintained in the absence of activity⁸, implying that wiring specificity relies on the matching of recognition molecules on the surface of sensory and motor neurons. However, determinants of fine synaptic specificity here, as in most regions of the central nervous system, have yet to be defined. To address the origins of synaptic specificity in these reflex circuits we have used molecular genetic methods to manipulate recognition proteins expressed by subsets of sensory and motor neurons. We show here that a recognition system involving expression of the class 3 semaphorin *Sema3e* by selected motor neuron pools, and its high-affinity receptor plexin D1 (*Plxnd1*) by proprioceptive sensory neurons, is a critical determinant of synaptic specificity in sensory–motor circuits in mice. Changing the profile of *Sema3e*–*Plxnd1* signalling in sensory or motor neurons results in functional and anatomical rewiring of monosynaptic connections, but does not alter motor pool specificity. Our findings indicate that patterns of monosynaptic connectivity in this prototypic central nervous system circuit are constructed through a recognition program based on repellent signalling.

Several observations led us to focus on the potential contribution of the class 3 semaphorin *Sema3e* and its receptors as mediators of sensory–motor synaptic specificity. *Sema3e* is expressed by a restricted set of brachial motor neurons and can serve as a bifunctional ligand, eliciting repellent responses through engagement of *Plxnd1* and attractant responses through interactions with a neuropilin 1 (*Nrp1*)–*Plxnd1* receptor complex^{9–12}. Moreover, *Sema3e* expression is lost, and the pattern of monosynaptic sensory–motor connections altered, in mice mutant for *Pea3* (also called *Etv4*), an ETS transcription factor expressed by several brachial motor neuron pools^{11,13}.

We analysed the role of *Sema3e* and its receptors in two sensory–motor reflex arcs. In one reflex arc that supplies the triceps forelimb muscle, motor neurons receive monosynaptic input from triceps sensory afferents¹³. In a second, atypical, reflex arc that controls the cutaneous maximus muscle, motor neurons fail to receive monosynaptic input from cutaneous maximus afferents^{13–15}. Cutaneous maximus motor neurons also lack monosynaptic input from triceps proprioceptive neurons, or from any other proprioceptive afferents, and conversely, triceps motor neurons lack monosynaptic input from cutaneous maximus proprioceptive afferents¹³. The contrasting

circuitry of these two reflex arcs permitted us to examine monosynaptic connectivity between proprioceptive sensory and motor neurons, as well as pool specificity, revealed by the absence of connections between functionally unrelated sensory–motor pairs.

To define the brachial motor neurons that express *Sema3e*, we identified motor neuron pools by retrograde labelling in *Sema3e*^{nlz} mice in which nuclear LacZ is expressed from the *Sema3e* locus (Fig. 1a and Supplementary Figs 1 and 2a). In p0 *Sema3e*^{nlz/+} mice, all LacZ-positive (*LacZ*^{on}) motor neurons were confined to the cutaneous maximus pool and ~85% of all cutaneous maximus motor neurons expressed LacZ (Fig. 1a, b). In contrast, neither *Pea3* nor LacZ was expressed in triceps motor neurons (Fig. 1a and Supplementary Fig. 2b)¹³. We also determined the profile of expression of *Sema3e* receptors by sensory neurons. At embryonic day (e)16.5, *Nrp1* expression was excluded from *TrkC*^{on} (also called *Ntrk3*) proprioceptive neurons (Supplementary Fig. 3a–d). In contrast, ~90% of *Plxnd1*^{on} dorsal root ganglion (DRG) neurons co-expressed *TrkC*, although only ~50% of all *TrkC*^{on} DRG neurons expressed *Plxnd1* (Fig. 1c and Supplementary Fig. 3g). *Plxnd1* was expressed by ~80% of *TrkC*^{on} proprioceptive neurons labelled retrogradely from the cutaneous maximus muscle nerve, by ~50% of *TrkC*^{on} triceps proprioceptors, but by only ~5% of *TrkC*^{on} ulnar proprioceptors (Fig. 1b, d, data not shown). Thus, within the cutaneous maximus reflex arc, which lacks monosynaptic connectivity, most motor neurons express *Sema3e* and most proprioceptive sensory neurons express *Plxnd1*, whereas in the monosynaptically connected triceps reflex arc the expression of *Plxnd1* by proprioceptors is not accompanied by motor neuron *Sema3e* expression (Fig. 1b).

We used gene targeting to assess the contribution of *Sema3e* and *Plxnd1* to the formation and specificity of monosynaptic connections in the cutaneous maximus and triceps reflex arcs. We analysed homozygous null *Sema3e*^{nlz/nlz} mice (Supplementary Fig. 1b–d) and conditional *Plxnd1*^{fllox} mice in which *Plxnd1* function in DRG neurons was eliminated by intercrossing with *Wnt1*^{Cre} mice^{16,17}. *Sema3e*^{nlz/nlz} and *Plxnd1*^{fllox/+} *Wnt1*^{Cre} (*Plxnd1*^{cond}) mice are viable and fertile, permitting studies of sensory–motor connectivity at post-natal stages. Analysis of *LacZ*^{on}, f-dextran-labelled motor neurons in p1 *Sema3e*^{nlz/+} and *Sema3e*^{nlz/nlz} mice revealed no difference in the positions of the cell bodies and dendrites of cutaneous maximus or triceps motor neurons (Supplementary Fig. 4). *Pea3* mutant mice show pronounced defects in cutaneous maximus motor neuron positioning and dendrite patterning^{11,13}, but our findings indicate that *Sema3e* does not mediate these aspects of *Pea3*'s influence on motor neuron differentiation.

To assess the contribution of *Sema3e* signalling to the functional specificity of monosynaptic connections in sensory–motor reflex arcs we stimulated, separately, the cutaneous maximus and triceps muscle

¹Biozentrum, Department of Cell Biology, University of Basel, 4056 Basel, Switzerland, and Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland. ²Howard Hughes Medical Institute, Departments of Neuroscience and Biochemistry and Molecular Biophysics, Kavli Institute for Brain Science, Columbia University, New York, USA. ³Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA.

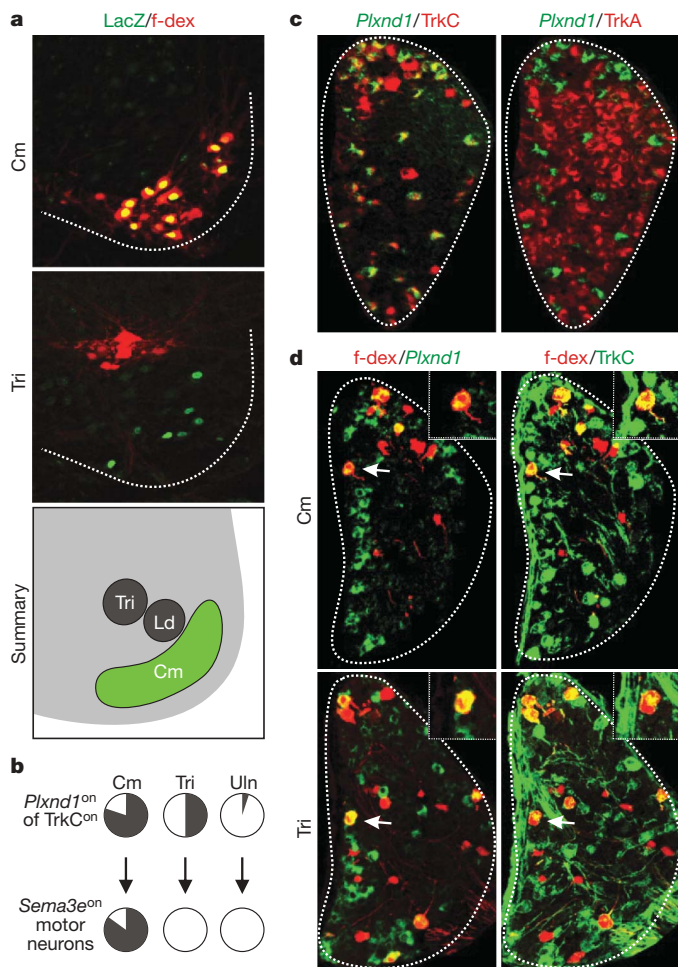


Figure 1 | *Sema3e* marks cutaneous maximus motor neurons, and *Plxnd1* marks proprioceptive neurons. **a**, *Sema3e* directed LacZ in p0 *Sema3e*^{nlz} spinal cords after retrograde f-dextran tracing from cutaneous maximus (Cm) or triceps (Tri) muscle nerves and summary of cutaneous maximus, latissimus dorsi (Ld) and triceps motor pool position. **b**, *Plxnd1* expression in TrkC^{on} proprioceptive afferents (top) and *Sema3e* expression in motor neurons (bottom) projecting to cutaneous maximus, triceps or ulnar (Uln) muscles. **c**, *Plxnd1* expression in e16.5 cervical level (c7) DRG, matched with TrkC and TrkA (also called Ntrk1). **d**, *Plxnd1* (left) and TrkC (right) expression in e16.5 DRG neurons after retrograde f-dextran tracing from cutaneous maximus or triceps muscle nerves.

nerves in spinal cord preparations isolated from p5 to p7 mice and recorded intracellular responses from identified motor neurons (Supplementary Figs 5 and 6)¹³. In wild-type mice, cutaneous maximus motor neurons lacked monosynaptic input after stimulation of cutaneous maximus afferents (Fig. 2a)¹³. In contrast in *Sema3e*^{nlz/nlz} mice, 45% of cutaneous maximus motor neurons received monosynaptic input after stimulation of cutaneous maximus afferents (mean latency 2.9 ± 0.2 ms; 21 out of 47 cells in 10 mice; Fig. 2b). Cutaneous maximus motor neurons in *Sema3e*^{nlz/nlz} mice still lacked monosynaptic input from triceps afferents (mean latency 8.8 ± 0.8 ms; $n = 1$ out of 35 cells in 6 mice; Fig. 2h). Moreover, the pattern of monosynaptic connectivity in the triceps arc was unaltered in *Sema3e*^{nlz/nlz} mice: triceps motor neurons received monosynaptic input from triceps but not cutaneous maximus sensory afferents (Fig. 2e, g, h). We conclude that the loss of *Sema3e* expression by cutaneous maximus motor neurons permits monosynaptic input from cutaneous maximus but not from triceps sensory afferents. The loss of *Plxnd1* expression from proprioceptive neurons elicited a similar change in the pattern of monosynaptic connectivity within the cutaneous maximus and triceps reflex arcs. In *Plxnd1*^{cond} mice, 43% of cutaneous maximus motor neurons received monosynaptic input

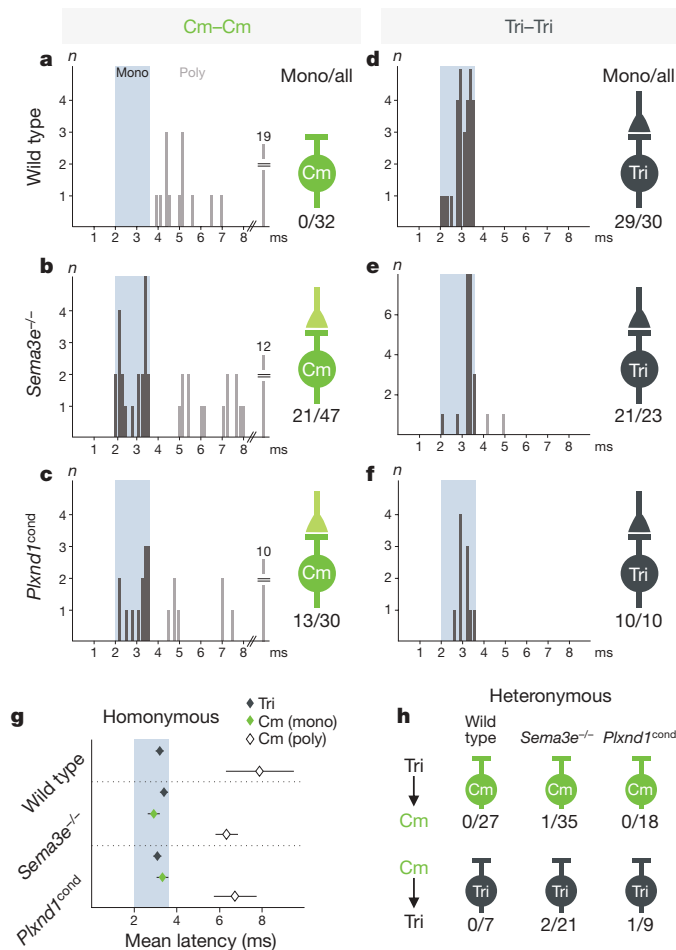


Figure 2 | Loss of *Sema3e* or *Plxnd1* function perturbs monosynaptic specificity. **a–f**, Cutaneous maximus and triceps onset latency histograms upon homonymous muscle nerve stimulation (0–8 ms in 0.15-ms bins). The blue box indicates monosynaptic latency. The number of neurons with onset latencies >8 ms is shown to the right of the x-axis break in **a–c**. A summary diagram showing the number of neurons with monosynaptic input/number of cells assayed is positioned to the right of each histogram. *Sema3e* or *Plxnd1* mutation leads to monosynaptic Cm–Cm connections (light green), absent in wild type (no presynaptic terminal). **g**, Mean onset latencies (ms \pm s.e.m.) of homonymous pairs. For cutaneous maximus, data within (green) and outside (white) the monosynaptic window are shown. There are significant differences between cutaneous maximus and triceps in wild-type mice, and between cutaneous maximus (poly) and triceps and cutaneous maximus (poly) and cutaneous maximus (mono) in both *Sema3e*^{−/−} and *Plxnd1*^{cond} mice ($P \leq 0.001$; Student's *t*-test). **h**, Summary of heteronymous recordings (Tri–Cm and Cm–Tri).

from cutaneous maximus sensory afferents (mean onset latency 3.3 ± 0.2 ms; 13 out of 30 cells in 7 mice; Fig. 2c, g), but they still lacked monosynaptic input from triceps sensory afferents (Fig. 2h). Triceps motor neurons received monosynaptic input from triceps but not cutaneous maximus sensory afferents in *Plxnd1*^{cond} mice (Fig. 2f–h). These findings provide functional evidence that *Sema3e* and *Plxnd1* interact during the formation of proprioceptive sensory–motor reflex circuits.

The functional changes in cutaneous maximus connectivity observed in *Sema3e*^{nlz/nlz} and *Plxnd1*^{cond} mice were accompanied by a localized reorganization of proprioceptive sensory terminals, revealed by analysis of the pattern of expression of the vesicular glutamate transporter vGlut1 (ref. 18). In wild-type mice, vGlut1^{on} proprioceptive terminals are distributed sparsely within the cutaneous maximus domain (Fig. 3a and Supplementary Figs 7 and 8), and anatomically identified cutaneous maximus motor neurons were contacted only by very few proprioceptive terminals (Fig. 3c, d). In

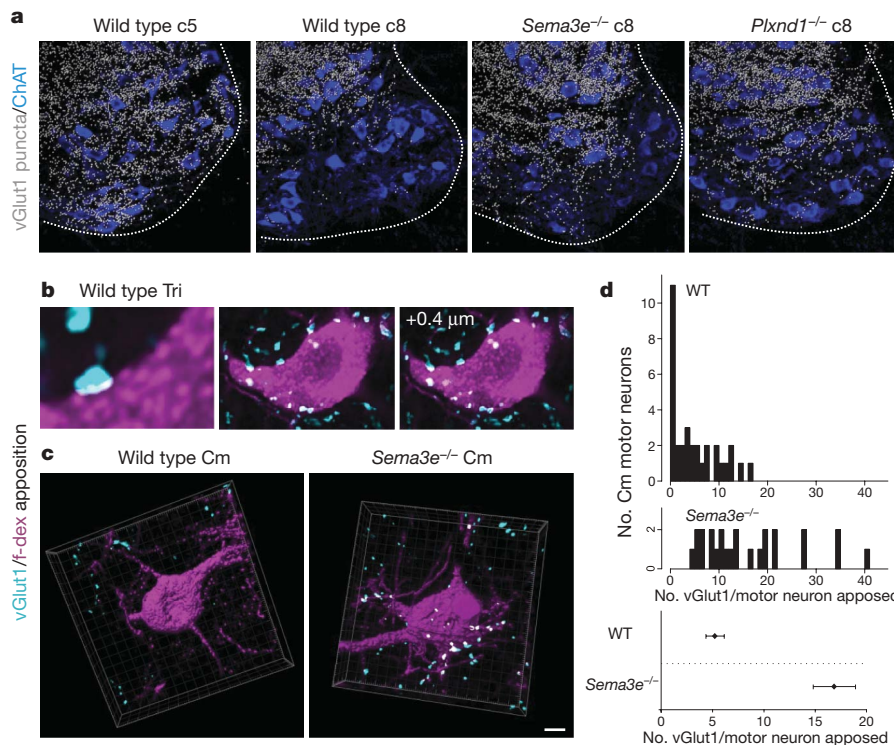


Figure 3 | Increase in vGlut1 sensory contacts with cutaneous maximus motor neurons in *Sema3e* mutant mice. **a**, Digitalized vGlut1^{on} proprioceptive terminals overlaid with ChAT^{on} motor neuron cell bodies in wild-type (cervical level (c5 and c8), *Sema3e*^{-/-} (c8) and *Plxnd1*^{cond} (c8) mice. **b**, **c**, vGlut1 apposition with f-dex-labelled triceps (**b**) or cutaneous maximus (**c**) motor neurons in wild-type (**b**, **c**) and *Sema3e*^{-/-} (**c**) mice.

Single confocal planes (**b**) or three-dimensional view (**c**) are shown. White indicates contact site. **d**, Frequency histograms of vGlut1 input to cutaneous maximus motor neurons. The bottom panel shows average vGlut1 number in contact with cutaneous maximus motor neurons (\pm s.e.m., $P \leq 0.001$; Student's *t*-test). Scale bar: **a**, 30 μ m; **c**, 10 μ m.

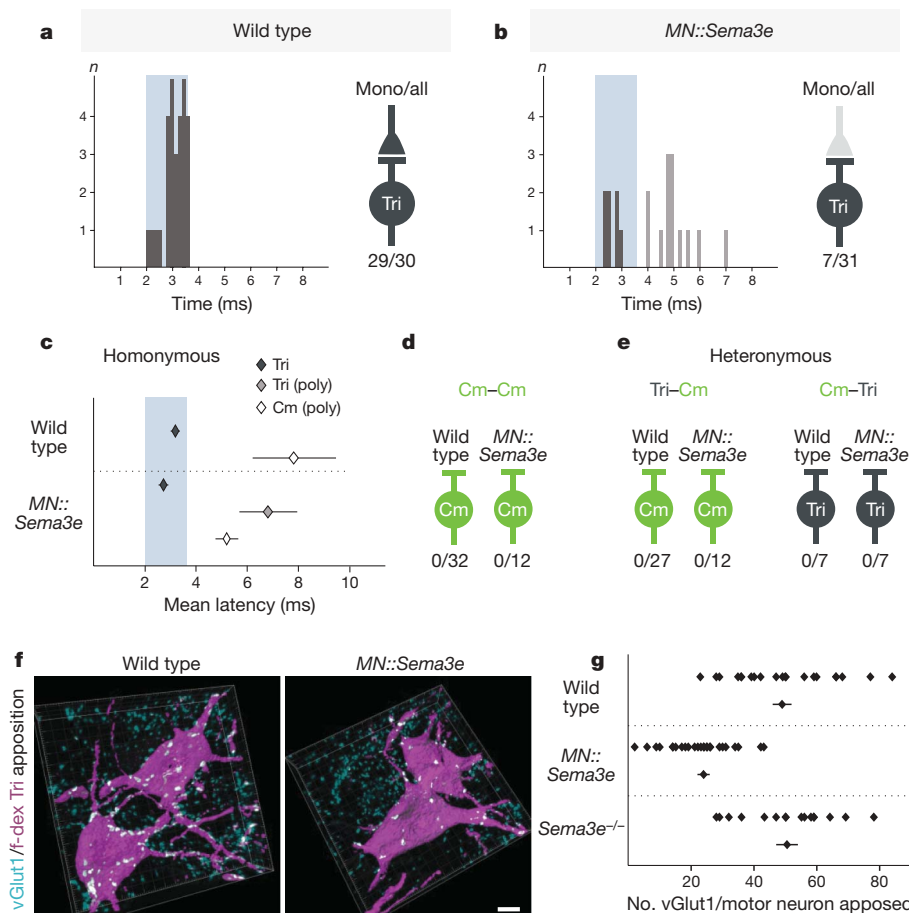


Figure 4 | Ectopic expression of *Sema3e* in motor neurons prevents monosynaptic connectivity. **a**, **b**, Triceps onset latency histograms upon triceps muscle nerve stimulation (0–8 ms in 0.15-ms bins). The blue box indicates monosynaptic latency. A summary diagram showing number of neurons with monosynaptic input/number of cells assayed is shown to the right of each histogram. The light-grey colour of the presynaptic terminal in **b** indicates reduction in triceps motor neurons with monosynaptic triceps input. **c**, Mean onset latencies (ms \pm s.e.m.) measured from homonymous pairs. For triceps, data within (dark grey) and outside (light grey) monosynaptic window is indicated. There are significant differences between cutaneous maximus and triceps in wild-type mice ($P \leq 0.001$; Student's *t*-test) and between triceps (mono) and triceps (poly) ($P \leq 0.01$; Student's *t*-test) and triceps (mono) and cutaneous maximus ($P \leq 0.001$; Student's *t*-test) in *MN::Sema3e* mice. **d**, **e**, Summary of Cm–Cm, Tri–Cm and Cm–Tri recordings. **f**, **g**, Analysis of vGlut1 apposition with f-dex-labelled triceps motor neurons (contact sites are white). **g**, The top rows show values from individual motor neurons; bottom rows show mean of analysed cells (\pm s.e.m.). There are significant differences between wild-type and *MN::Sema3e* mice and *Sema3e*^{-/-} and *MN::Sema3e* mice ($P \leq 0.001$; Student's *t*-test). Scale bar, 10 μ m.

contrast, in the cutaneous maximus domain of *Sema3e*^{nlz/nlz} and *Plxnd1*^{cond} mice we detected a ~5-fold increase in the density of vGlut1^{on} proprioceptive terminals (Fig. 3a and Supplementary Fig. 8) and a ~3-fold increase in the number of proprioceptive synaptic contacts with identified cutaneous maximus motor neurons (Fig. 3d). Thus, monosynaptic sensory input to cutaneous maximus motor neurons in the absence of *Sema3e*–*Plxnd1* signalling seems to result from the formation of new proprioceptive sensory contacts with cutaneous maximus motor neurons.

We also examined whether *Sema3e* signalling by motor neurons is sufficient to repel proprioceptive synaptic inputs. To test this, we directed *Sema3e* expression to triceps (and all other spinal) motor neurons by crossing a Cre-inducible *Tau*^{*Sema3e.ires.nlz*} mouse line with a *ChAT*^{cre} line, to generate *MN::Sema3e* mice (Supplementary Fig. 9). In *MN::Sema3e* mice, >95% of brachial motor neurons now

expressed *Sema3e* at levels equal or greater than that of endogenous *Sema3e* (Supplementary Fig. 9b–d). In *MN::Sema3e* mice, only 23% of triceps motor neurons received monosynaptic input from triceps afferents, compared with >95% in wild-type mice (Fig. 4a–c and Supplementary Fig. 10). Yet, triceps and cutaneous maximus motor neurons still lacked monosynaptic input from cutaneous maximus afferents (Fig. 4d, e). Furthermore, we detected a ~50% reduction in the number of vGlut1^{on} terminal contacts with the cell bodies and proximal dendrites of triceps motor neurons in *MN::Sema3e* mice compared with wild-type mice (Fig. 4f, g). The persistence of some vGlut1^{on} afferent terminal contacts with triceps motor neurons that express *Sema3e*, and the corresponding preservation of physiological inputs, may have its basis in the expression of *Plxnd1* by only about one-half of all triceps proprioceptive neurons, and could also reflect the existence of synaptic inputs from *Plxnd1*^{off} proprioceptive neurons that supply synergistic muscles. We conclude that *Sema3e* expression by triceps motor neurons is sufficient to reduce the incidence of monosynaptic input from triceps sensory afferents.

Our findings show that the specificity of monosynaptic sensory–motor connections in the mammalian spinal cord depends on a recognition system in which the matching expression of a *Sema* ligand and its plexin receptor prevents synapse formation. Yet even for these relatively simple reflex circuits, an additional layer of specificity is evident (Fig. 5a). Elimination of *Sema3e*–*Plxnd1* signalling uncovers a latent propensity of cutaneous maximus sensory afferents to form monosynaptic connections with cutaneous maximus motor neurons, but triceps motor neurons remain offlimits (Fig. 5a). Similarly, triceps sensory afferents stripped of *Plxnd1* still ignore cutaneous maximus motor neurons. These observations indicate that a recognition system independent of *Sema3e*–*Plxnd1* signalling underlies motor pool specificity (Fig. 5a). Other *Sema* and plexin proteins are expressed by motor and sensory neurons^{12,16} and thus pool specificity could involve a more elaborate matrix of *Sema*–plexin recognition. More generally, our findings extend the influence of repellent *Sema* signalling from the guidance and long-range pruning of axons^{19,20} (see Supplementary Information) to fine synaptic specificity.

The recruitment of a repellent recognition system to the cause of synaptic specificity in spinal sensory–motor circuits provides an intriguing contrast with recent studies of connectivity in the nematode motor system and vertebrate retina, where specificity seems to depend on adhesive interactions between immunoglobulin superfamily proteins expressed by pre- and post-synaptic partners^{21,22}. Whether the intricate patterns of sensory–motor connectivity in the mammalian spinal cord emerge solely through layers of repellent filtering or also involve adhesive recognition remains to be determined. Nevertheless, a prominent role for repellent recognition in synaptic specificity seems likely, given recent evidence that Wnt-mediated repellent signalling controls the position and selectivity of synaptic contacts in invertebrate neural circuits^{23,24}.

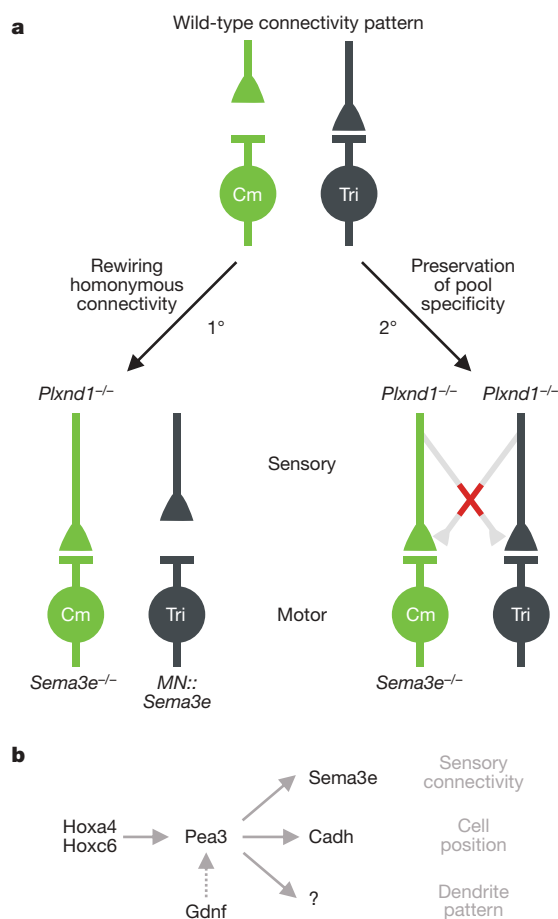


Figure 5 | Regulation of synaptic specificity by repellent *Sema3e*–*Plxnd1* signalling. **a**, Connectivity patterns in cutaneous maximus and triceps reflex arcs in wild-type mice (top) and in mice with altered *Sema3e*–*Plxnd1* signalling (bottom). Top: in wild-type mice, triceps but not cutaneous maximus motor neurons receive direct homonymous proprioceptive inputs. Bottom left: changing the profile of *Sema3e*–*Plxnd1* expression rewires homonymous connectivity. Loss of *Sema3e*–*Plxnd1* signalling results in monosynaptic connections between cutaneous maximus afferents and cutaneous maximus motor neurons; ectopic *Sema3e* expression reduces monosynaptic connections between triceps afferents and motor neurons. Bottom right: changing *Sema3e*–*Plxnd1* signalling does not erode pool specificity: no aberrant heteronymous sensory–motor connections are observed (red cross). **b**, Genetic program for *Sema3e* expression in cutaneous maximus motor neurons. Hox transcription factors program cutaneous maximus motor neuron pool identity²⁸, conferring cutaneous maximus motor neuron sensitivity to the retrograde inductive influence of Gdnf²⁹. Pea3 expression by cutaneous maximus motor neurons controls cell position, dendrite pattern and the selectivity of proprioceptive inputs through distinct downstream pathways^{11,13,30}.

METHODS SUMMARY

Sema3e^{nlz} mice were generated using a targeting strategy similar to that described¹⁰ but by the integration of a *creERT2-IRES-Nls-lacZ-pA* cassette into the endogenous start codon of the *Sema3e* locus. For the generation of *Tau*^{*Sema3e*} mice, a *lox-STOP-lox-Sema3e-IRES-Nls-lacZ-pA* targeting cassette was integrated into exon 2 of the *Tau* locus^{25,26}. *Plxnd1*^{fllox} mice were generated by insertion of *loxP* sites flanking the first coding exon of the *Plxnd1* locus as described¹⁷. Electrophysiological recordings from p5 to p7 animals and analysis were carried out as described¹³. Retrograde tracing experiments from cutaneous maximus and triceps muscles were carried out by f-dextran injection at p12 and followed by analysis of vGlut1 input to f-dextran-labelled motor neurons at p14. Cryostat sections were processed for immunohistochemistry by sequential application of primary antibodies and fluorophore-conjugated secondary antibodies (Invitrogen and Jackson Laboratories)²⁷. For *in situ* hybridization experiments on cryostat sections, *Plxnd1* (accession number BC019530; NCBI) and *Sema3e*¹¹ plasmids were used to generate probes. Images were collected on an Olympus confocal or dissection microscope. For quantitative analysis of synaptic input to cutaneous maximus and triceps motor neurons, the IMARIS co-localization tool

(version 6.1.5.) was used to assess proprioceptive input contacting motor neurons retrogradely labelled by f-dextran, using images acquired at 0.2 μ m confocal steps and $\times 150$ magnification.

Received 17 December 2008; accepted 18 March 2009.

Published online 6 May 2009.

- Eccles, J. C., Eccles, R. M. & Lundberg, A. The convergence of monosynaptic excitatory afferents on many different species of alpha motoneurons. *J. Physiol.* **137**, 22–50 (1957).
- Rossignol, S., Dubuc, R. & Gossard, J. P. Dynamic sensorimotor interactions in locomotion. *Physiol. Rev.* **86**, 89–154 (2006).
- Windhorst, U. Muscle proprioceptive feedback and spinal networks. *Brain Res. Bull.* **73**, 155–202 (2007).
- Brown, A. *Organization in the Spinal Cord* (Springer, 1981).
- Burke, R. E. & Glenn, L. L. Horseradish peroxidase study of the spatial and electrotonic distribution of group Ia synapses on type-identified ankle extensor motoneurons in the cat. *J. Comp. Neurol.* **372**, 465–485 (1996).
- Mears, S. C. & Frank, E. Formation of specific monosynaptic connections between muscle spindle afferents and motoneurons in the mouse. *J. Neurosci.* **17**, 3128–3135 (1997).
- Baldissera, F., Hultborn, H. & Illert, M. *Handbook of Physiology: Integration in Spinal Neuronal Systems* 509–595 (American Physiological Society, 1981).
- Frank, E. The formation of specific synaptic connections between muscle sensory and motor neurons in the absence of coordinated patterns of muscle activity. *J. Neurosci.* **10**, 2250–2260 (1990).
- Chauvet, S. *et al.* Gating of *Sema3E*/PlexinD1 signaling by neuropilin-1 switches axonal repulsion to attraction during brain development. *Neuron* **56**, 807–822 (2007).
- Gu, C. *et al.* Semaphorin 3E and plexin-D1 control vascular pattern independently of neuropilins. *Science* **307**, 265–268 (2005).
- Livet, J. *et al.* ETS gene *Pea3* controls the central position and terminal arborization of specific motor neuron pools. *Neuron* **35**, 877–892 (2002).
- Cohen, S. *et al.* A semaphorin code defines subpopulations of spinal motor neurons during mouse development. *Eur. J. Neurosci.* **21**, 1767–1776 (2005).
- Vrieseling, E. & Arber, S. Target-induced transcriptional control of dendritic patterning and connectivity in motor neurons by the ETS gene *Pea3*. *Cell* **127**, 1439–1452 (2006).
- Theriault, E. & Diamond, J. Intrinsic organization of the rat cutaneous trunci motor nucleus. *J. Neurophysiol.* **60**, 463–477 (1988).
- Holstege, G., van Neerven, J. & Evertse, F. Spinal cord location of the motoneurons innervating the abdominal, cutaneous maximus, latissimus dorsi and longissimus dorsi muscles in the cat. *Exp. Brain Res.* **67**, 179–194 (1987).
- Yoshida, Y., Han, B., Mendelsohn, M. & Jessell, T. M. PlexinA1 signaling directs the segregation of proprioceptive sensory axons in the developing spinal cord. *Neuron* **52**, 775–788 (2006).
- Zhang, Y. *et al.* *Tie2*Cre-mediated inactivation of *plexinD1* results in congenital heart, vascular and skeletal defects. *Dev. Biol.* **325**, 82–93 (2009).
- Oliveira, A. L. *et al.* Cellular localization of three vesicular glutamate transporter mRNAs and proteins in rat spinal cord and dorsal root ganglia. *Synapse* **50**, 117–129 (2003).
- Bagri, A., Cheng, H. J., Yaron, A., Pleasure, S. J. & Tessier-Lavigne, M. Stereotyped pruning of long hippocampal axon branches triggered by retraction inducers of the semaphorin family. *Cell* **113**, 285–299 (2003).
- Low, L. K., Liu, X. B., Faulkner, R. L., Coble, J. & Cheng, H. J. Plexin signaling selectively regulates the stereotyped pruning of corticospinal axons from visual cortex. *Proc. Natl Acad. Sci. USA* **105**, 8136–8141 (2008).
- Shen, K., Fetter, R. D. & Bargmann, C. I. Synaptic specificity is generated by the synaptic guipost protein SYG-2 and its receptor, SYG-1. *Cell* **116**, 869–881 (2004).
- Yamagata, M. & Sanes, J. R. Dscam and Sidekick proteins direct lamina-specific synaptic connections in vertebrate retina. *Nature* **451**, 465–469 (2008).
- Inaki, M., Yoshikawa, S., Thomas, J. B., Aburatani, H. & Nose, A. Wnt4 is a local repulsive cue that determines synaptic target specificity. *Curr. Biol.* **17**, 1574–1579 (2007).
- Klassen, M. P. & Shen, K. Wnt signaling positions neuromuscular connectivity by inhibiting synapse formation in *C. elegans*. *Cell* **130**, 704–716 (2007).
- Hippenmeyer, S. *et al.* A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol.* **3**, e159 (2005).
- Kramer, I. *et al.* A role for Runx transcription factor signaling in dorsal root ganglion sensory neuron diversification. *Neuron* **49**, 379–393 (2006).
- Arber, S., Ladle, D. R., Lin, J. H., Frank, E. & Jessell, T. M. ETS gene *Er81* controls the formation of functional connections between group Ia sensory afferents and motor neurons. *Cell* **101**, 485–498 (2000).
- Dasen, J. S., Tice, B. C., Brenner-Morton, S. & Jessell, T. M. A Hox regulatory network establishes motor neuron pool identity and target-muscle connectivity. *Cell* **123**, 477–491 (2005).
- Haase, G. *et al.* GDNF acts through PEA3 to regulate cell body positioning and muscle innervation of specific motor neuron pools. *Neuron* **35**, 893–905 (2002).
- Price, S. R., De Marco Garcia, N. V., Ranscht, B. & Jessell, T. M. Regulation of motor neuron pool sorting by differential expression of type II cadherins. *Cell* **109**, 205–216 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Ladle for help in analysis of electrophysiological data, J. Livet and C. Henderson for discussions, and P. Schwab, A. Ponti and M. Stadler for help in image and statistical analysis. M. Mendelsohn, J. Kirkland and B. Han helped in the generation of *Plxnd1^{cond}* mice and J. F. Spetz, P. Kopp and B. Kuchemann in the generation of *Sema3e* mutant and *MN::Sema3e* mice. R. Axel, P. Caroni, E. Frank, A. Hantman, C. Henderson, D. Ladle and A. Luethi provided comments on the manuscript. E.P.-V., M.S. and S.A. are supported by the Swiss National Science Foundation, NCCR Frontiers in Genetics, the Kanton Basel-Stadt, EU Framework Program 7 and the Novartis Research Foundation. Y.Y. is supported by NIH grant RO1NS065048. T.M.J. is an HHMI Investigator, and is supported by grants from NINDS, EU Framework Program 7, Project ALS, The Harold and Leila Mathers Foundation, and The Wellcome Trust.

Author Contributions E.P.-V., M.S. and Y.Y. made critical primary contributions to this study. E.P.-V. performed physiological analysis of sensory–motor connectivity and tracing experiments in wild-type and mutant mice. M.S. participated in *Sema3e* and *Plxnd1* subpopulation assignment, anatomical analyses of sensory–motor organization and generated *Sema3e^{niz/niz}* and *MN::Sema3e* mice. Y.Y. found that *Plxnd1* is expressed by subsets of proprioceptive sensory neurons and generated *Plxnd1^{cond}* mice. T.M.J. and S.A. initiated complementary aspects of this project, analysed data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.A. (silvia.arber@unibas.ch) or T.M.J. (tmj1@columbia.edu).

Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger

Gang G. Wang¹, Jikui Song², Zhanxin Wang², Holger L. Dormann¹, Fabio Casadio¹, Haitao Li², Jun-Li Luo³, Dinshaw J. Patel² & C. David Allis¹

Histone H3 lysine 4 methylation (H3K4me) has been proposed as a critical component in regulating gene expression, epigenetic states, and cellular identities¹. The biological meaning of H3K4me is interpreted by conserved modules including plant homeodomain (PHD) fingers that recognize varied H3K4me states^{1,2}. The dysregulation of PHD fingers has been implicated in several human diseases, including cancers and immune or neurological disorders³. Here we report that fusing an H3K4-trimethylation (H3K4me3)-binding PHD finger, such as the carboxy-terminal PHD finger of PHF23 or JARID1A (also known as KDM5A or RBBP2), to a common fusion partner nucleoporin-98 (NUP98) as identified in human leukaemias^{4,5}, generated potent oncoproteins that arrested haematopoietic differentiation and induced acute myeloid leukaemia in murine models. In these processes, a PHD finger that specifically recognizes H3K4me3/2 marks was essential for leukaemogenesis. Mutations in PHD fingers that abrogated H3K4me3 binding also abolished leukaemic transformation. NUP98–PHD fusion prevented the differentiation-associated removal of H3K4me3 at many loci encoding lineage-specific transcription factors (*Hox(s)*, *Gata3*, *Meis1*, *Eya1* and *Pbx1*), and enforced their active gene transcription in murine haematopoietic stem/progenitor cells. Mechanistically, NUP98–PHD fusions act as ‘chromatin boundary factors’, dominating over polycomb-mediated gene silencing to ‘lock’ developmentally critical loci into an active chromatin state (H3K4me3 with induced histone acetylation), a state that defined leukaemia stem cells. Collectively, our studies represent, to our knowledge, the first report that deregulation of the PHD finger, an ‘effector’ of specific histone modification, perturbs the epigenetic dynamics on developmentally critical loci, catastrophizes cellular fate decision-making, and even causes oncogenesis during mammalian development.

Recent studies have shown that an H3K4me3-binding PHD finger in the human BPTF (also known as NURF301), ING2 or TFIID (also known as TBP) complex helps to recruit and/or stabilize these effectors and associated factors onto appropriate target promoters during transcriptional regulation^{1,6–10}. A non-methylated H3K4 (H3K4me0)-engaging PHD finger in the DNMT3L or LSD1 (also known as AOF2) complex connects the activities of DNA methylation or H3K4 demethylation to repressive chromatin^{11,12}. Notably, germ-line mutation in the PHD finger of RAG2 abrogates its recognition of H3K4me3 and causes immunodeficiency¹³. Mutations in the PHD finger of ING1 have been implicated in cancers^{3,8,14}. However, evidence supporting a causal role for PHD finger mutation and inappropriate interpretation of histone modification in oncogenesis is still elusive.

In clinically reported acute myeloid leukaemia (AML) patients^{4,5}, chromosomal translocation fuses the C-terminal PHD finger of JARID1A or PHF23 (JARID1A–PHD3 or PHF23–PHD), together with

nuclear localization signals, to NUP98, a common leukaemia fusion partner that contains transactivation activities^{15–17} (Supplementary Fig. 1). Notably, the JARID1A–PHD3 motif is excluded from an alternatively spliced isoform of JARID1A and the corresponding NUP98–JARID1A fusion (hereafter referred to as NJS), whereas it is retained in the longer fusion isoform (hereafter referred to as NJL; Fig. 1a). We asked whether JARID1A–PHD3 as a putative chromatin-‘reading’ module is involved in haematopoietic malignancies. To test this, we examined the leukaemogenic potential of both fusion isoforms using a haematopoietic progenitor transformation assay¹⁸ (Supplementary Fig. 2a). Murine bone-marrow-derived haematopoietic stem/progenitor cells transduced with an empty retrovirus or a retrovirus encoding NJS proliferated transiently and differentiated into mature cells, whereas cells transduced with NJL proliferated indefinitely as undifferentiated progenitors (Fig. 1b, c). The NJL-transduced marrow cells proliferated in a cell-autonomous manner, had typical myeloblast morphology (Fig. 1d) and expressed early myeloid progenitor antigens (c-Kit⁺ Cd11b⁺ Cd34⁺ Gr-1[–] Cd19[–] B220^{–/low}; Fig. 1e and Supplementary Fig. 2b). The arrest of myeloid differentiation by NJL indicated that it would induce leukaemia *in vivo*. Indeed, all of the 12 mice transplanted with murine marrow progenitors transduced with NJL died of AML in an average of 69 days, whereas those reconstituted with empty vector- or NJS-transduced progenitors remained healthy after 1 year (Fig. 1f). NJL-induced leukaemia exhibited a myeloid phenotype (Supplementary Fig. 2c, d), and typically presented with an enlarged spleen, packed progenitors in the bone marrow, and a massive increase in peripheral white blood cells (Supplementary Table 1 and Fig. 1g, h). Taken together, NJL represents a potent leukaemia oncogene both in cellular and animal models.

The fact that NJS failed to induce leukaemia indicated that the PHD finger is required for leukaemogenesis. Indeed, deletion of JARID1A–PHD3, but not of JARID1A sequences before or after it, abolished NJL-mediated transformation of haematopoietic cells (Supplementary Fig. 2f–h). We next addressed whether JARID1A–PHD3 recognizes histone methylation. First, only histone H3 associated with recombinant JARID1A–PHD3 using total histone extracts (Supplementary Fig. 3a). When a mini-library of H3 peptides containing unmodified, mono-, di- or tri-methylated Lys 4, Lys 9, Lys 27, Lys 36 or Lys 79 was screened in a biotinylated peptide pull-down assay, JARID1A–PHD3 only interacted with those containing H3K4me3/2 (Fig. 2a and Supplementary Fig. 3b). Such specificity was further confirmed by immunostaining and co-immunoprecipitation using Flag–NJL stable expression cells: NJL exhibited a speckled nuclear staining pattern, and significantly co-localized with H3K4me3 but not with H3K9me3 (Supplementary Fig. 4). Most of the NJL was bound to mononucleosomes containing H3K4me3 but not H3K27me3 (Supplementary Fig. 3c and Fig. 2d). Calorimetry-based measurements

¹Laboratory of Chromatin Biology & Epigenetics, The Rockefeller University, New York, New York 10065, USA. ²Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ³Department of Cancer Biology, The Scripps Research Institute, Scripps Florida, Jupiter, Florida 33458, USA.

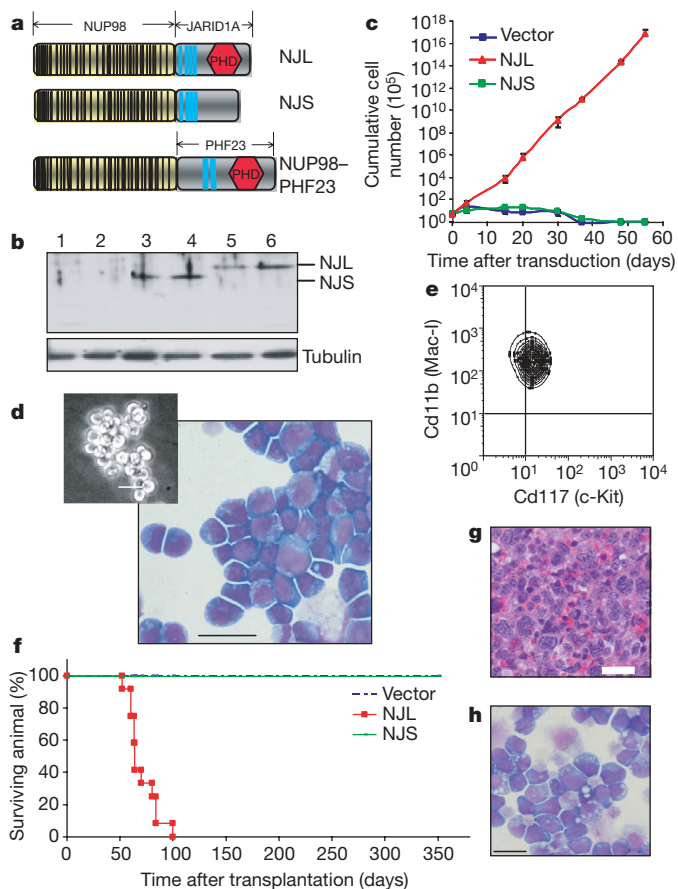


Figure 1 | The PHD finger-containing NUP98-JARID1A fusion isoform (NJL), but not that lacking the PHD finger (NJS), confers leukaemogenic potentials to haematopoietic stem/progenitor cells. **a**, The NUP98-JARID1A and NUP98-PHF23 structures are shown (see Supplementary Fig. 1 for details). **b**, Immunoblot of haematopoietic cells transduced with empty vector (lanes 1–2) or that encoding Flag-tagged NJS (lanes 3–4) or NJL (lanes 5–6). **c**, Proliferation kinetics of lineage-negative haematopoietic cells after transduction of empty vector, NJL or NJS. Data are presented as mean \pm s.d. of six experiments. **d**, **e**, Wright-Giemsa staining (**d**, insert, microscopy image) and fluorescence-activated cell sorting analysis (FACS, **e**) of NJL-transformed cells. **f**, Leukaemia kinetics in mice (12 each group) after transplantation of bone marrow transduced with vector, NJL or NJS. **g**, Haematoxylin and eosin staining of spleen section and **h**, Wright-Giemsa staining of bone marrow from NJL-induced AML mice. Scale bars, 20 μ m.

revealed a dissociation constant (K_d) of $\sim 0.75 \mu$ M for JARID1A-PHD3 binding to H3K4me3, with reduced affinities for binding to H3K4me2/1/0 (Supplementary Fig. 3d).

We determined the structure of JARID1A-PHD3-H3K4me3 complexes using X-ray crystallographic and NMR spectroscopic techniques. Both analyses showed that the JARID1A-PHD3-H3K4me3 interaction was established by (1) anti-parallel β -sheet pairing between the H3 backbone and a β -sheet of JARID1A-PHD3; (2) a hydrophobic cleft formed by two Trp residues (Trp 1625 and Trp 1635) that anchor the H3K4me3 side chain; and (3) the positioning of H3R2 in an acidic pocket (Glu 1627/Asp 1629/Asp 1633) (Fig. 2b and Supplementary Figs 5b and 6c). H3K4me3 is stacked between the indole rings of two orthogonally aligned Trp residues, with intermolecular contacts shown in Fig. 2b and Supplementary Figs 5b and 6d. The X-ray (a domain-swapped dimer of one molecule and a crystallographic symmetry-related molecule) and solution NMR (monomer) analyses are summarized in Supplementary Fig. 5 (statistics in Supplementary Table 2) and Supplementary Figs 6 and 7 (statistics in Supplementary Table 3), respectively. Comparison between JARID1A-PHD3 structures in the free and H3K4me3-bound

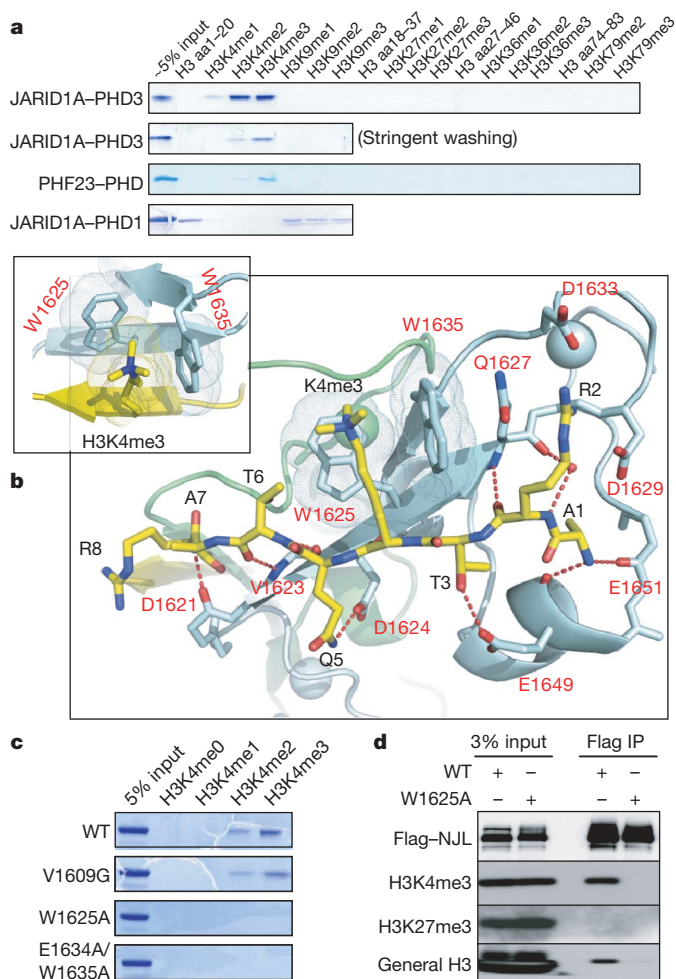


Figure 2 | JARID1A-PHD3, an essential motif for NJL-mediated leukaemia, specifically recognizes H3K4me3/2 marks. **a**, Capability of JARID1A-PHD3, PHF23-PHD and JARID1A-PHD1 (the first PHD finger of JARID1A; Supplementary Fig. 1) to interact with H3 peptides containing different states of Lys methylation, in a peptide pull-down assay. JARID1A-PHD1 interacted with H3K4me0 as BHC80-PHD¹¹. aa, amino acids. **b**, The crystal structure of JARID1A-PHD3 (cyan) complexed with H3K4me3 peptide (yellow), and a close-up view of the H3K4me3-binding channel (inset) formed by two orthogonally aligned Trp residues. The numeration of JARID1A-PHD3 and H3 residues is shown in red and black, respectively. Protein Data Bank accession number, 3GL6. **c**, Capability of wild-type (WT) or mutant JARID1A-PHD3 to bind to H3K4me3/2. **d**, Co-immunoprecipitation showing that NJL containing the wild-type, but not mutant (W1625A) PHD finger, associated with H3K4me3 or H3 in transiently transfected 293 cells.

state (Supplementary Fig. 6a, b) showed no overall conformational changes. Residues Trp 1625 and Trp 1635 are evolutionarily conserved among JARID1 homologues (Supplementary Fig. 8a). Mutations targeting these Trp residues disrupted the H3K4me3-binding both *in vitro* (Fig. 2c) and in cells (Fig. 2d). Such a two-sided H3K4me3-binding tryptophan channel is a variant form of the H3K4me3-engaging pocket involving 3–4 hydrophobic residues found in the PHD finger of BPTF⁷, ING2 (ref. 8), *Saccharomyces cerevisiae* Yng1 (ref. 19) or RAG2 (ref. 13) (Supplementary Fig. 8b–d). Yet, it exhibited a stronger H3K4me3-binding affinity ($K_d = 0.75 \mu$ M). Collectively, the PHD finger, an essential motif of NUP98-JARID1A, uniquely recognizes H3K4me3/2 using an aromatic engaging channel.

To gain insight into the mechanisms of NJL-induced AML, we used microarray analyses to compare the transcriptome of NJL-transformed marrow progenitors and that of control murine cells, committed myeloid progenitors generated as described before¹⁸. Notably, a considerable portion of genes upregulated in the NJL-transformed

progenitors were those either targeted by polycomb proteins^{20,21} or exhibiting a 'bivalent domain pattern'²² in stem cells, many of which encode developmentally critical transcription factors (*Hoxa5/a7/a9/a10*, *Gata3*, *Meis1*, *Eya1* and *Pbx1*; Supplementary Table 4). Such upregulation was further confirmed by RT-PCR using vector- versus NJL-transduced marrow cells (Supplementary Fig. 9a–c). Other *Hoxa* genes (*a1*, *a2*, *a11* and *a13*) were not expressed in NJL-transformed progenitors. We detected a similar target specificity for *Hoxa* genes using chromatin immunoprecipitation (ChIP)—NJL directly bound to the promoters of *Hoxa6–a10*, but not to distal *Hoxa1–a3* or *Hoxa11–a13* (Fig. 3a and Supplementary Fig. 9d, green). NJL-binding specificity among *Hox* clusters was correlated to H3K4me3, as H3K4me3 was abundant in *Hoxa6–a10*, but was low/absent in *Hoxa1–a4* or *Hoxa11–a13* (Fig. 3b). Enforced expression of *Hox* and *Meis1* has been shown sufficient to induce AML²³. This indicated that NJL blocks haematopoietic differentiation and induces AML by enforcing the transcription of these genes.

It has been reported that A-cluster *Hox* gene expression is high in haematopoietic stem cells and early progenitors, followed by down-regulation and shut-off during terminal differentiation²⁴. Our *ex vivo* murine haematopoietic stem/progenitor cell system recapitulated such dynamics: coincident to the silencing of a haematopoietic stem cell marker and the activation of a differentiation marker (Supplementary Fig. 9f), *Hoxa9* and *Hoxa10* were downregulated >10- or 60-fold, respectively, in 8 days of culture (Fig. 3c). The concurrent loss of *Hoxa9*- and *Hoxa10*-associated H3K4me3 was observed in these cells (Fig. 3e). Notably, NJL persistently enforced high levels of *Hoxa9/a10* expression and *Hoxa9/a10*-associated H3K4me3 in marrow cells, whereas *Hoxa9/a10* expression were silenced 10 days after transduction of vector or NJS in similarly maintained cells (Fig. 3c–e). To rigorously test the role of H3K4me3 recognition during leukaemogenesis, we mutated the H3K4me3-engaging residues. NJL containing a mutation on Trp 1625 or Trp 1635 failed to bind to H3K4me3 or H3 (Fig. 2d), failed to bind to the *Hoxa9* promoter that exhibited high H3K4me3 in 293 cells (Fig. 4a and

Supplementary Fig. 9i), failed to enforce the *Hoxa9* expression (Fig. 4b) or *Hoxa9*-associated H3K4me3 in haematopoietic progenitors (Fig. 4c), and failed to transform the haematopoietic cells (Fig. 4d), whereas the irrelevant mutation Val1609Gly did not affect these activities (Supplementary Fig. 10e). To assess whether the NJL-induced phenotype was unique to JARID1A-PHD3, we investigated another similar *de novo* translocation, NUP98–PHF23 (Fig. 1a)⁵, and also swapped JARID1A-PHD3 with other PHD fingers reported before. PHF23-PHD specifically engaged H3K4me3/2 as predicted (Fig. 2a)¹; NUP98–PHF23 robustly enforced *Hoxa9*-associated H3K4me3 and transformed haematopoietic progenitors (Fig. 4c, e and Supplementary Fig. 10). Notably, swapping JARID1A-PHD3 with another H3K4me3/2-binding PHD finger from ING2 (ref. 8) or even *S. cerevisiae* Yng1 (ref. 19) also succeeded in the transformation, whereas replacing it with an H3K4me0-binding PHD finger, either BHC80-PHD¹¹ or JARID1A-PHD1 (Fig. 2a), abolished the transformation (Fig. 4c, e). Therefore, engaging H3K4me3/2 by NUP98–PHD fusion causes leukaemia by enforcing an active state on developmentally critical loci.

Because the H3K4me3 recognition cannot provide DNA sequence specificity and yet NJL-upregulated genes were enriched with polycomb-targeted^{20,21} or 'bivalent domain' genes²² in stem cells (for example, *Hox* genes, *Gata3* and *Meis1*; Supplementary Table 4), we asked whether such specificity is due to their dynamically regulated characteristics. Towards this end, we examined the effect of NJL on two distinct gene classes: developmentally critical genes, and house-keeping genes that exhibit constitutive H3K4me3 (Supplementary Fig. 11a, top panel). Interestingly, although NJL bound to housekeeping genes, it had little effect on their expression during cell differentiation (Supplementary Fig. 11a, middle and bottom panels). Thus, NJL tends to affect the developmentally critical loci specifically during haematopoiesis. We next pursued the possibility that NJL interferes with the activities of polycomb proteins at these developmentally critical loci. Using ChIP, we found that whereas Ezh2 or Suz12 was spread throughout *Hoxa* clusters in vehicle-infected marrow progeni-

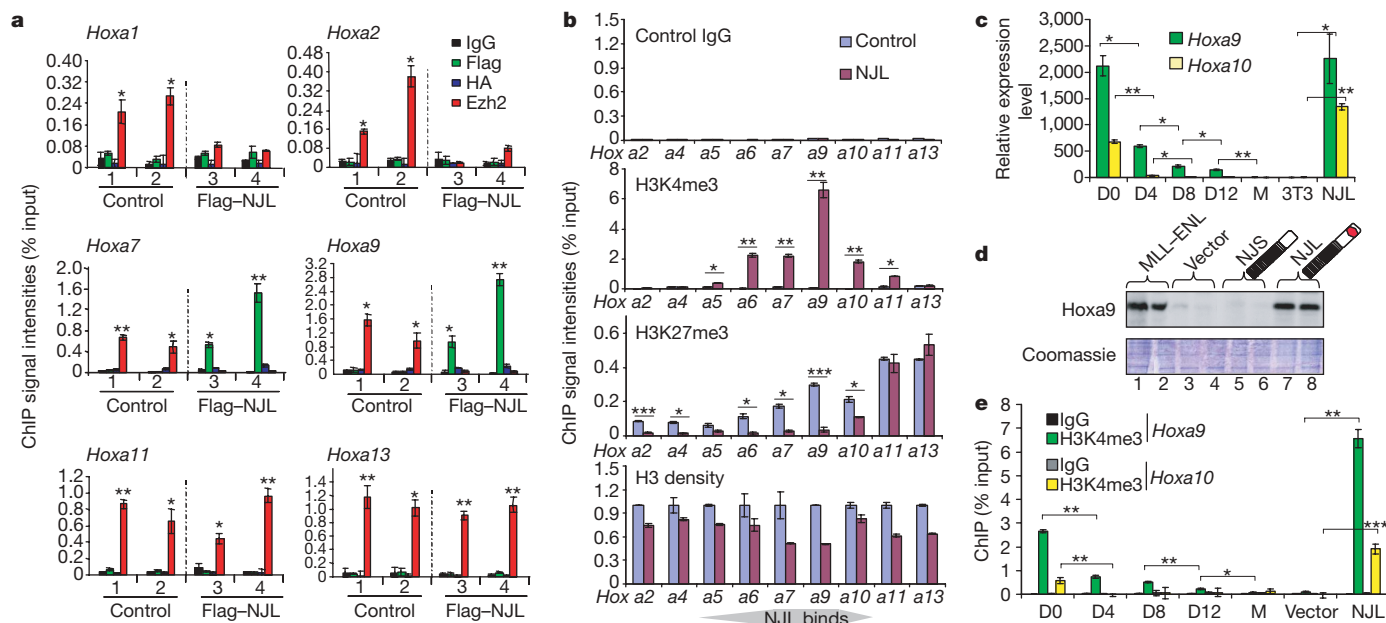


Figure 3 | NUP98–JARID1A enforced high H3K4me3 and active transcription associated with developmentally critical loci such as *Hox*. **a**, ChIP for NJL- or Ezh2-binding to *Hoxa* gene promoters in a committed myeloid progenitor line¹⁸ (1), or in haematopoietic stem/progenitor cells 3 weeks after transduction of control vector (2), or 3×Flag-tagged NJL (3–4). HA, haemagglutinin. **b**, ChIP of H3K4me3, H3K27me3 and general H3 among the *Hoxa* gene cluster in haematopoietic progenitors 3 weeks after transduction of vector or NJL. **c**, *Hoxa9* and *Hoxa10* expression in

haematopoietic stem/progenitor cells after *in vitro* cultivation (0, 4, 8 and 12 days (D)), and in macrophages (M), NIH-3T3 fibroblasts or NJL-transformed progenitors. **d**, Anti-Hoxa9 blot in marrow progenitors 10 days after transduction of MLL-ENL, empty vector, NJS or NJL. **e**, ChIP for *Hoxa9* or *Hoxa10* promoter-associated H3K4me3 in haematopoietic stem/progenitor cells after days of *in vitro* culture, and in macrophages and marrow progenitors 20 days after transduction of vector or NJL. $n = 3$, error bars indicate s.d.; * $P < 0.01$, ** $P < 0.001$ and *** $P < 10^{-4}$.

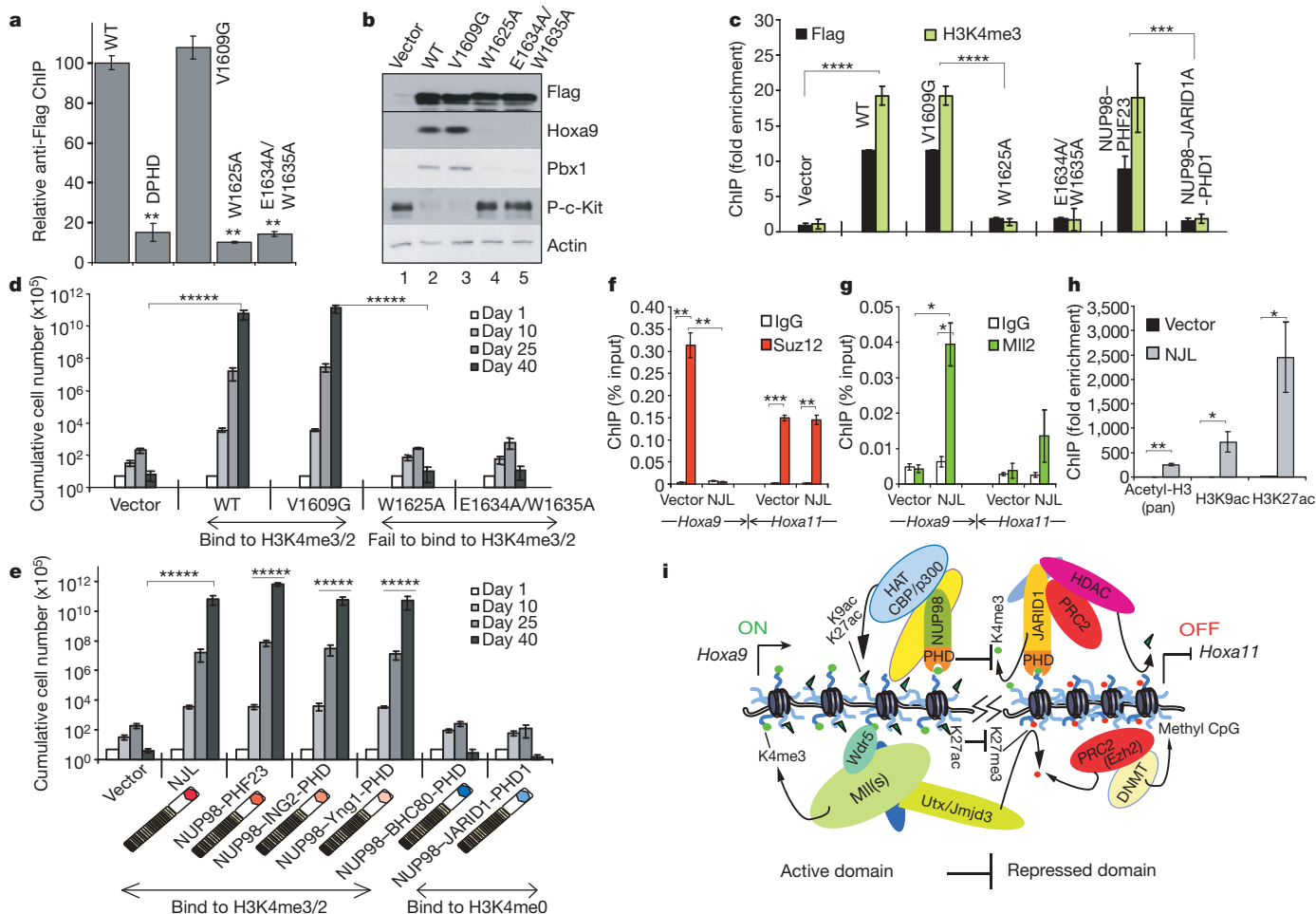


Figure 4 | The H3K4me3/2 engagement by NUP98–JARID1A perturbs the epigenetic state of developmentally critical loci during haematopoiesis.

a, The impact of mutations on Flag–NJL binding to *HOXA9* in 293 cells. WT, wild type. **b**, Immunoblot of haematopoietic progenitors 10 days after transduction of vector, wild-type or mutant NJL. Phosphorylated c-Kit (P-c-Kit) is a marker of mast cells. Actin is shown as a loading control. **c**, ChIP for *Hoxa9* promoter-associated NUP98–fusion proteins (3×Flag-tagged) and H3K4me3 in marrow progenitors 10 days after transduction. **d**, **e**, Transforming capacities after introducing mutation to NJL (**d**) or those by NUP98–PHF23 (**e**) or after replacing JARID1A–PHD3 with another PHD

tors that underwent differentiation, these polycomb factors were restricted within *Hoxa11–a13* in NJL-infected progenitors (Figs 3a and 4f, red). In the NJL-transduced cells, H3K27me3 was also only detected at *Hoxa13–a11*; the differentiation-associated spreading of H3K27me3 was inhibited at a region from *Hoxa10* to *Hoxa1* (Fig. 3b). The spreading of polycomb factors from distal *Hox* loci (*a13–a11*) seemed to be blocked at *Hoxa10–a9* by NJL that was bound there (Fig. 3a and Supplementary Fig. 9d). A similar result was also found at *Meis1* (Supplementary Fig. 9e). Consistent with previous reports^{15,16}, the recruitment of p300 (also known as EP300) and a marked increase in H3 acetylation (>2,000-fold for H3K27 acetylation) were observed on *Hoxa9* in NJL-transduced cells (Fig. 4h and Supplementary Fig. 11b). Collectively, NUP98–PHD fusion dominated over the spreading of polycomb and enforced an H3K4me3/acetylated histone state at developmentally critical loci, an epigenetic state that defines leukaemia stem cells.

We have demonstrated for the first time, to our knowledge, that fusing an H3K4me3-engaging PHD finger (plus the nuclear localization signal) to a common partner such as NUP98 is sufficient to induce leukaemia. We showed that NUP98–PHD fusion prevented the silencing of critical loci encoding master transcription factors (*Hox* genes, *Gata3*, *Meis1* and *Pbx1*) during haematopoietic differentiation.

finger that engages either H3K4me3/2 or H3K4me0. The total progenitor number was counted at day 1, 10, 25 and 40. **f–h** ChIP for Suz12 (**f**), Mll2-binding to *Hoxa9/a11* (**g**), and *Hoxa9*-associated H3 acetylation (**h**) in marrow progenitors 15 days after transduction of vector or NJL. Error bar indicates s.d.; $n = 3$; $*P < 0.05$, $**P < 0.005$, $***P < 10^{-4}$ and $****P < 10^{-6}$. **i**, A scheme showing that NUP98–PHD fusion acts as a boundary factor and prevents the spreading of polycomb factors from *Hoxa13/a11* to *Hoxa9*, thus inhibiting H3K4me3 removal and H3K27me3 addition during haematopoiesis.

NUP98 fusion partners can be placed into two groups: DNA-binding homeodomain and chromatin-associated factors including PHD fingers (JARID1A and PHF23)¹⁷. Although the existence of an additional unknown ligand is possible for PHD fingers in the latter group (as the H3K4 site cannot be mutated in mammals), the most straightforward interpretation of our findings is that binding H3K4me3/2 marks are responsible for leukaemia described here. In support, a genetic interaction was demonstrated in yeast between H3K4 and the Yng1 PHD finger²⁵, a module that imparted similar oncogenic properties when swapped into our assays (Fig. 4e). Several PHD fingers exist in NSD1, another NUP98–fusion partner¹⁶, however, none contains critical H3K4me3-engaging residues¹. Thus, our report demonstrates that inappropriate interpretation of histone modification can actively induce a deregulation of developmentally critical loci, perturb cellular/epigenetic identities, and even induce oncogenesis. NUP98–PHD fusion coordinates acts of H3K4me3/2 and histone acetylation, mimicking mechanisms used by evolutionarily conserved ING(s)-complexes for robust gene activation^{19,26} (Supplementary Fig. 12). H3K4me3 bound by NUP98–PHD may serve as a ‘seed’ of propagation mediated by Wdr5–Mll2/3 complexes^{1,27} that is also coupled with Utx/Jmjd3-mediated H3K27 demethylation^{28,29}, as we detected high levels of Wdr5, Rbbp5 and Mll2 on *Hoxa9* in

NJL-transduced marrow cells (Fig. 4g and Supplementary Fig. 11c, d). We suggest that NUP98-PHD acts as a 'boundary factor', using the PHD finger to protect H3K4me3 from Jarid1-mediated demethylation²⁹ and also inducing H3K27ac to block H3K27me addition (Fig. 4i). In support, we observed a bivalent domain feature²² at *Hoxa11-a10*, the junction region of two antagonizing mechanisms (Fig. 3b). A loss-of-function mutation of RAG2-PHD in immunodeficiency, and a gain-of-function mutation involving PHD fingers in malignancies described here, indicate new types of disease that arise from 'misinterpreting' the 'histone code'^{33,30}. With ~200 PHD fingers in the human genome and some intimately associated with diseases³, we expect similar 'mis-reading' mechanisms may be responsible for some unstudied diseases. These pathologies, together with those caused by 'mis-writing' or 'mis-erasing'²⁹ histone modifications, underscore the importance in investigating the biological readout of histone marks.

METHODS SUMMARY

Haematopoietic cell transformation assays. Protocols for the culture of primary murine haematopoietic stem/progenitor cells were previously described¹⁸. In brief, 100,000 lineage-negative bone marrow stem/progenitor cells were subjected to retroviral infection, followed by kinetics analyses of proliferation versus differentiation in *ex vivo* culture system as described before¹⁸.

Peptide pull-down assay. Pull-down using biotinylated histone peptide and recombinant protein was performed as described^{6,11}. After binding, peptide-avidin beads were washed extensively in solution containing 50 mM Tris, pH 7.5, 150 mM NaCl (250 mM as stringent washing), 0.05% NP-40, 0.3 mg ml⁻¹ BSA and 1 mM dithiothreitol (DTT).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 November 2008; accepted 2 April 2009.

Published online 10 May 2009; corrected 11 June 2009 (see full-text HTML version for details).

- Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell* **25**, 15–30 (2007).
- Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Struct. Mol. Biol.* **14**, 1025–1040 (2007).
- Baker, L. A., Allis, C. D. & Wang, G. G. PHD fingers in human diseases: Disorders arising from misinterpreting epigenetic marks. *Mutat. Res.* **647**, 3–12 (2008).
- van Zutven, L. J. *et al.* Identification of NUP98 abnormalities in acute leukemia: JARID1A (12p13) as a new partner gene. *Genes Chromosom. Cancer* **45**, 437–446 (2006).
- Reader, J. C., Meekins, J. S., Gojo, I. & Ning, Y. A novel NUP98-PHF23 fusion resulting from a cryptic translocation t(11;17)(p15;p13) in acute myeloid leukemia. *Leukemia* **21**, 842–844 (2007).
- Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86–90 (2006).
- Li, H. *et al.* Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91–95 (2006).
- Pena, P. V. *et al.* Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**, 100–103 (2006).
- Shi, X. *et al.* ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**, 96–99 (2006).
- Vermeulen, M. *et al.* Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58–69 (2007).
- Lan, F. *et al.* Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* **448**, 718–722 (2007).
- Ooi, S. K. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* **448**, 714–717 (2007).
- Matthews, A. G. *et al.* RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* **450**, 1106–1110 (2007).
- Gong, W., Suzuki, K., Russell, M. & Riabowol, K. Function of the ING family of PHD proteins in cancer. *Int. J. Biochem. Cell Biol.* **37**, 1054–1065 (2005).

- Kasper, L. H. *et al.* CREB binding protein interacts with nucleoporin-specific FG repeats that activate transcription and mediate NUP98-HOXA9 oncogenicity. *Mol. Cell. Biol.* **19**, 764–776 (1999).
- Wang, G. G., Cai, L., Pasillas, M. P. & Kamps, M. P. NUP98-NSD1 links H3K36 methylation to *Hox-A* gene activation and leukaemogenesis. *Nature Cell Biol.* **9**, 804–812 (2007).
- Moore, M. A. *et al.* NUP98 dysregulation in myeloid leukemogenesis. *Ann. NY Acad. Sci.* **1106**, 114–142 (2007).
- Wang, G. G., Pasillas, M. P. & Kamps, M. P. Meis1 programs transcription of *FLT3* and cancer stem cell character, using a mechanism that requires interaction with Pbx and a novel function of the Meis1 C-terminus. *Blood* **106**, 254–264 (2005).
- Taverna, S. D. *et al.* Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. *Mol. Cell* **24**, 785–796 (2006).
- Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
- Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353 (2006).
- Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
- Kroon, E. *et al.* *Hoxa9* transforms primary bone marrow cells through specific collaboration with *Meis1a* but not *Pbx1b*. *EMBO J.* **17**, 3714–3725 (1998).
- Pineault, N., Helgason, C. D., Lawrence, H. J. & Humphries, R. K. Differential expression of *Hox*, *Meis1*, and *Pbx1* genes in primitive cells throughout murine hematopoietic ontogeny. *Exp. Hematol.* **30**, 49–57 (2002).
- Martin, D. G. *et al.* The Yng1p plant homeodomain finger is a methyl-histone binding module that recognizes lysine 4-methylated histone H3. *Mol. Cell. Biol.* **26**, 7871–7879 (2006).
- Doyon, Y. *et al.* ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol. Cell* **21**, 51–64 (2006).
- Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859–872 (2005).
- Lee, M. G. *et al.* Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* **318**, 447–450 (2007).
- Cloos, P. A., Christensen, J., Agger, K. & Helin, K. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev.* **22**, 1115–1140 (2008).
- Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Baker, P. Chi, A. Ruthenberg and A. Xiao for critical reading and help in manuscript preparation, C. Hughes for antibodies, and other Allis laboratory members for their advice. We are extremely grateful to M. Kamps and S. Rafii for shared plasmids and expert advice. Thanks to E. Coutavas, W. Chen, E. Ezhkova and the R. Roeder laboratory for help on cell and animal work. G.G.W. is supported by a Leukemia & Lymphoma Society Fellow award and a Choh-Hao Li Memorial Fund Scholar award; H.L.D. is supported by a predoctoral Boehringer Ingelheim Foundation fellowship; This research was supported by the National Institutes of Health (NIH) grant and funds of Rockefeller University to C.D.A., funds of Abby Rockefeller Mauze Trust and the Dewitt Wallace and Maloris Foundations to D.J.P., US Department of Defense CDMRP grant to J.-L.L., and a joint Starr Foundation Cancer Consortium grant.

Author Contributions G.G.W. and C.D.A. designed the study. G.G.W. performed most of the cellular and molecular experiments, and wrote the paper; J.S. performed protein preparation, NMR structure determination and crystallization; Z.W. performed crystallographic analyses; H.L.D. and G.G.W. did the immunostaining; F.C. participated in plasmid/protein preparation; H.L. performed isothermal titration calorimetry measurement; G.G.W. and J.-L.L. performed animal studies; D.J.P. and C.D.A. supervised the structural and functional aspects of the project, respectively, and helped with manuscript preparation.

Author Information The structural coordinates of JARID1A-PHD3 in the H3-bound or free state have been submitted to the Protein Data Bank under accessions 3GL6, 2KGG and 2KGI, and the chemical shift assignment of NMR structures to BioMagResBank under accessions 16209 and 16210. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.D.A. (alliscd@rockefeller.edu).

METHODS

Plasmid construction and retroviral expression system. The human *NUP98-JARID1A* fusion complementary DNA⁴ was generated by ligating *NUP98* sequences encoding amino acids 1–514 to those encoding amino acids 1489–1690 of *JARID1A* transcript variant 1 (NCBI accession number NM_001042603) or amino acids 1489–1641 of *JARID1A* transcript variant 2 (NCBI accession number NM_005056), producing two fusion isoforms (NJL or NJS), respectively. The same method was used to generate *NUP98-PHF23* (ref. 5). The fusion cDNA with an amino-terminal 3×FLAG was cloned into MSCV retroviral expression vector (Clontech). *JARID1A*, *PHF23* and *BHC80* cDNAs were purchased from Open Biosystems. *NUP98* plasmids were provided by J. M. van Deursen, *Hoxa9* by M. P. Kamps, *MLL-ENL* by R. K. Slany, *Yng1* by S. D. Taverna, *CBX7* by E. Bernstein, and *ING2* by Z. Tang. Mutations were generated by site-directed mutagenesis, and all used plasmids were confirmed by sequencing.

Purification and culture of primary haematopoietic cells. Bone marrow cells collected from the femur and tibia of balb/c or b/6 mice were subject to lineage-negative (Lin^-) enrichment using Haematopoietic Progenitor Enrichment Kit (StemCell Technologies or Miltenyi Biotec) to remove cells expressing differentiation antigens as described before¹⁶. Approximately 400,000 Lin^- -enriched haematopoietic progenitors were obtained per mouse with ~10% c-Kit⁺ Lin^- Sca1⁺ haematopoietic stem cells. Before retroviral infection, Lin^- -enriched haematopoietic progenitors were stimulated in OptiMEM base medium (Invitrogen) complemented with 10% FBS (Invitrogen), 1% antibiotics, 50 μM β -mercaptoethanol and a cytokine cocktail containing stem-cell factor (SCF; supernatant of SCF-producer cells), 5 ng ml⁻¹ FLT3 ligand (Sigma), 5 ng ml⁻² IL3 and IL6 (Miltenyi) for 2–3 days as described^{18,31}. After retroviral infection and selection (1 $\mu\text{g ml}^{-1}$ puromycin), marrow cells were plated in the same medium with SCF as the sole cytokine. Cell splitting and replating to fresh medium were performed every 3–4 days to keep the cell number <2 million per well (6- or 12-well plates). Cell morphology was examined by Wright–Giemsa staining. Macrophages were obtained by culture of marrow cells in M-CSF (Miltenyi) for 1–2 weeks as described³². Immortalized cell lines that mimic committed neutrophil-macrophage progenitors were generated as described previously^{18,31,33}.

Murine bone marrow transplantation leukaemogenic assay. The leukaemogenic potentials of oncogenes were evaluated in sublethally irradiated syngeneic mice, after tail vein injection with 100,000 bone-marrow-derived Lin^- cells that were infected with retrovirus encoding the fusion gene as described¹⁸. Mice exhibiting leukaemic phenotype were subjected to pathological analyses.

Recombinant protein production and glutathione S-transferase pull-down. *JARID1A-PHD3* (amino acids 1601–1660) glutathione S-transferase (GST)-fusion proteins were produced using a previously described protocol¹⁹. GST pull-down using total histone extracts was performed as described with modification⁹. In brief, ~2 μg GST-fusion protein bound to glutathione beads (Amersham) were incubated with 10 μg of calf thymus histone extracts (Worthington) in a binding buffer containing 50 mM Tris-HCl, pH 7.5, 0.5 M NaCl, 0.5% NP-40, 0.2 mM EDTA, 1 mM DTT and protease inhibitor cocktail (Roche) at 4 °C for 4 h.

Native co-immunoprecipitation. Mononucleosome-containing fractions were prepared as described before⁶. In brief, intact nuclei were subject to limited micrococcal nuclease (MNase) digestion, so that the major form of released chromatin is mononucleosome. After the removal of the insoluble fraction by centrifugation, supernatant containing mononucleosomes was then incubated with Flag or HA-agarose beads (Sigma), or with Dynal magnetic beads (Invitrogen) coupled with anti-H3K4me3 (Abcam) or control antibodies. After extensive washing, precipitated proteins were subject to immunoblot.

Isothermal titration calorimetry measurements. Calorimetric experiments were conducted at 25 °C with a MicroCal iTC200 instrument as described⁷. Recombinant *JARID1A-PHD3* protein and H3_{1–15}K4me (H3 amino acids 1–15, with Lys4 methylated) peptides were dialysed overnight against 25 mM Tris-HCl, pH 7.5, 50 mM KCl and 2 mM β -mercaptoethanol. Protein concentration was determined by absorbance spectroscopy (Tyr $\epsilon_{280} = 1,420 \text{ M}^{-1} \text{ cm}^{-1}$; Trp $\epsilon_{280} = 5,600 \text{ M}^{-1} \text{ cm}^{-1}$; Cys $\epsilon_{280} = 125 \text{ M}^{-1} \text{ cm}^{-1}$). H3_{1–15}K4me peptides were quantified by the absorbance of an added C-terminal Tyr with $\epsilon_{280} = 1,280 \text{ M}^{-1} \text{ cm}^{-1}$ for peptide. Acquired calorimetric titration data were analysed using software Origin7.0 (MicroCal, iTC200) on the basis of a 1:1 binding stoichiometry.

Antibodies and immunoblot. Antibodies used were anti-Flag (Sigma; M2), anti-HA (Covance, MMS101), anti-Hoxa9 (Upstate, 07-178), anti-Pbx1 (Santa Cruz, sc889), anti-phospho-c-Kit (Cell Signaling) and anti-Tubulin (Sigma).

ChIP analysis. ChIP analysis was performed using an Upstate ChIP kit and a protocol described before³⁴. One-to-two-million cells per ChIP were used for histones, and 2–3 million for others. Antibodies and the amount used were anti-Flag (Sigma M2, 1–3 μg), anti-HA (Covance MMS101, 1–3 μg), anti-H3K4me3 (Upstate 07-473, 1 μl ; Abcam 8580, 0.5 μg), anti-H3K27me3 (Upstate 07-449, 0.5 μg), anti-acetyl-H3 (Upstate 06-599, 1 μg), anti-general H3 (Abcam

1791, 0.5 μg), anti-acetyl-H3K9 (Upstate 06-942, 1 μg), anti-acetyl-H3K27 (Abcam 4729, 1 μg), anti-Ezh2 (Cell Signaling 4905, 4–5 μl), anti-Suz12 (Upstate 07-379, 2 μl), anti-Mll2 (Bethyl A300-113A, 4 μg), anti-WDR5 (Upstate 07-706, 2 μg), anti-RBBP5 (Bethyl A300-109A, 3 μg ; a gift from C. Hughes) and anti-p300 (Santa Cruz, N15/C20, 10 μg). The same amount of nonspecific IgG (Upstate) was used as antibody control, and a silenced intragenic locus, Chr8Int, was used as a locus control for H3K4me3 or activator binding as described²¹. The promoter sequence was acquired from the UCSC genomic browser (<http://genome.ucsc.edu>). ChIP primers are shown in Supplementary Table 5. ChIP signals were represented as the percentage of signals from total chromatin used, and the fold of enrichment was calculated by normalizing against signals of nonspecific IgG.

Microarray analysis. Total RNA was extracted and the transcript expression was quantified using Affymetrix GeneChip Mouse arrays as described¹⁸. RNA hybridization, scanning and signal quantification were performed by the Rockefeller University Genomic Resources Center. Hybridization signals were retrieved and normalized, followed by differential expression analysis and statistical analysis using GeneSpring Analysis Platform GX 7.0 (Agilent Technologies).

RT-PCR analysis. Reverse transcription of RNA was performed using the random hexamer and Invitrogen Superscript III kit. Usually the PCR amplicon (~90–200 bp) is designed to span over large intron regions. Exon–intron information was obtained from the UCSC genomic browser. Quantitative PCR was performed in triplicate using SYBR green master mix reagent (Applied Biosystem) on a Stratagene Mx3005P QPCR system. Primer information is shown in Supplementary Table 5.

Flow cytometry. Cells were blocked with BD FcBlock (2.4G2) and stained on ice with fluoro-conjugated antibodies (1:1,000 dilution of Cd117-FITC, Sca-1-PE-CY7, Cd34-APC, Cd34-FITC, Cd11b-APC, Gr-1-PE, Cd19-PE or B220-PE, BD Biosciences), and analysed on a BD FACS Calibur cytometer. Data were collected and analysed using CellQuestPro and FlowJo software.

Immunofluorescence microscopy. Suspension-cultured haematopoietic cells were attached to coverslips treated with 0.01% (w/v) poly-lysine, followed by 15-min fixation in 4% paraformaldehyde and 10-min solubilization in PBS, 0.2% Triton-X100 and 0.2% NP-40. After a 30-min block in PBS, 2.5% BSA and 10% normal goat serum, cells were stained with primary antibodies (M2 anti-Flag (1:1,000–2,000 dilution of 1 mg ml⁻¹), rabbit anti-H3K4me3 (Upstate 07-473 or Abcam 8580, 1:2,000) or rabbit anti-H3K9me3 antibodies (Upstate 07-442, 1:1,000)) followed by washing and staining with fluorescent-labelled secondary antibodies. After washing, fluorescent signal was visualized and analysed with a DeltaVision Image Restoration Microscope (Applied Precision) and a Confocal Microscope (Olympus). Deconvolution microscopy image analysis was performed to reassign the out-of-focus blurred light to its origin³⁵, and subcellular co-localization analysis was carried out from stacks of deconvolved images using ImageJ (W. S. Rasband, <http://rsb.info.nih.gov/ij/>) and the plugin JACO³⁶. Confocal microscopy analysis was performed as previously described³⁷. Co-immunostaining statistics was analysed using Pearson's Coefficient of Correlation method. Image acquisition, processing and analyses were performed with help from Rockefeller University Bio-Imaging Center, and detailed protocols are available on request.

Statistics. All results are presented as the mean and s.d. Statistical analyses were performed using Student's *t*-test.

Protein preparation for structural studies. The gene fragment encoding *JARID1A-PHD3* was fused to the C terminus of a His(6×)-SUMO tag in a modified pRSFDuet-1 vector (Novagen), with a ubiquitin-like-protease (ULP) cleavage site located at the linker region. The bacterial expressed protein was purified using a Ni-NTA affinity column, followed by ULP cleavage, separation of *JARID1A-PHD3* from His(6×)-SUMO via a second Ni-NTA chromatography step, and gel filtration. The *JARID1A-PHD3*–H3K4me3 complex was obtained by mixing *JARID1A-PHD3* protein with an equal molar amount of H3_{1–9}K4me3 peptides (H3 amino acids 1–9, with Lys4 trimethylated), then purified by gel filtration, and concentrated by ultrafiltration.

Crystal growth. The crystals of *JARID1A-PHD3*–H3K4me3 complexes were obtained by equilibrating a reservoir consisting of 20% (w/v) poly(ethylene glycol) monomethyl ether 2000, 10 mM nickel (II) chloride hexahydrate, and 0.1 M Tris, pH 8.5, with a hanging drop consisting of 1 μl of the reservoir solution and 1 μl of a 27 mg ml⁻¹ protein solution in 10 mM Tris, pH 8.0, 0.1 mM ZnCl₂, 5 mM DTT and 50 mM NaCl (Crystal Screen 2 kit, Hampton Research). A mixture of the well solution with 10% (v/v) glycerol was used as a cryoprotectant.

Data collection and structure determination. An anomalous diffraction data set for the *JARID1A-PHD3*–H3K4me3 complex was collected at the zinc anomalous peak wavelength (1.28215 Å) at beamline NE-CAT 24ID-C, Advanced Photon Source, Chicago. The data set was indexed, integrated and merged to 2.2 Å using the program HKL2000. The crystal belongs to *I*₄ space group and contains one molecule per asymmetric unit. Heavy-atom search, single-wavelength anomalous dispersion (SAD) phasing and model building

were performed with the PHENIX³⁸ software package. Three zinc atoms were unambiguously identified for SAD phasing, and ~90% residues of the protein-peptide complex were successfully built into the initial model. The PHENIX-model was further manually rebuilt using COOT³⁹ and refined using REFMAC5⁴⁰ in successive cycles. The final refined structure has R_{work} and R_{free} values of 0.200 and 0.234, respectively (Supplementary Table 2). One molecule forms a domain-swapped dimer with a crystallographic symmetry-related molecule (Supplementary Fig. 5a). The swapped segment spans the first 14 residues from the N terminus.

Using a crystal of the complex following pH optimization of crystallization conditions, we were able to collect one 1.9 Å data set at wavelength 0.97949 Å at the same beamline. The crystal belongs to the same crystal form as the previous one. We solved the high-resolution structure by molecular replacement using PHASER⁴¹ with the above 2.2 Å model after removing all the water molecules and some flexible residues. Structure refinement was done using CNSolve⁴², cycled with manual model building in COOT. Before the refinement, the same R_{free} set of reflections were transferred from the low-resolution data using the program Freerflag in CCP4 suite⁴³ for effective cross validation. For both data sets, ~10% reflections were selected in a 'random' mode throughout the resolution range. After resetting the overall B-factor to 20 Å² and rigid body refinement, simulated annealing starting at 5,000K was performed to reduce model bias before extensive B-factor and positional refinement. The final model contains full-length JARID1A-PHD3 (1609–1659) with one extra serine at the N terminus from the expression vector, histone H3_{1–9}K4me3, three zinc ions and 32 water molecules. The JARID1A-PHD3–H3K4me3 complex in the crystal shows that one molecule forms a domain-swapped dimer with a crystallographically symmetry-related molecule (Supplementary Fig. 5a). Two zinc ions are integral to the folding of the PHD finger, whereas the third zinc ion locates at the interface between two domain-swapped dimers, thereby mediating crystal packing (Supplementary Fig. 5c). The R_{work} and R_{free} of the final structure are 0.208 and 0.225, respectively (Supplementary Table 2).

Isotopic labelling, NMR data collection and structure determination. Samples used for NMR chemical shift assignments, ¹⁵N relaxation measurements, and structure determination contained 0.2–0.5 mM uniformly-¹⁵N- or [¹³C, ¹⁵N]-labelled JARID1A-PHD3 in the free state and in complex with unlabelled H3_{1–9}K4me3 peptide dissolved in NMR buffer (20 mM Na-phosphate, 1 mM ZnCl₂, 5 mM DTT, 90% H₂O/10% D₂O) at pH 7.0. The sample used for measurements of ¹⁵N-¹H residual dipolar couplings (RDCs) contained 0.2 mM JARID1A-PHD3 aligned in 12 mg ml^{−1} of bacteriophage Pf1 (Alsa), 10 mM MOPS, 200 mM NaCl, pH 7.0.

All NMR spectra were collected at the New York Structural Biology Center using 800 MHz Bruker NMR spectrometers equipped with ¹H, ¹⁵N, ¹³C triple-resonance cryogenic probes. Unless indicated otherwise, the sample temperature was controlled at 20 °C. A suite of three-dimensional (3D) heteronuclear NMR experiments, including HNCACB, CBCA(CO)NH, HNCO, HBHA(CO)NH and HCCH-TOCSY were acquired for sequential backbone and non-aromatic side-chain assignments of JARID1A-PHD3 both in the free state and in complex with H3_{1–9}K4me3 peptide in solution. Two-dimensional (2D) nuclear Overhauser enhancement spectroscopy (NOESY) ($\tau_{\text{mix}} = 100$ ms), 3D ¹⁵N-edited NOESY-HSQC (heteronuclear single quantum correlation) ($\tau_{\text{mix}} = 100$ ms), 3D aromatic ¹³C-edited NOESY-HSQC ($\tau_{\text{mix}} = 100$ ms) and 3D aliphatic ¹³C-edited NOESY-HSQC ($\tau_{\text{mix}} = 100$ ms) data sets were acquired and used for additional assignments (side-chain amide and aromatic groups) and distance constraints. To selectively observe the nuclear Overhauser effect (NOEs) between JARID1A-PHD3 and H3_{1–9}K4me3 peptide, a [¹³C, ¹⁵N]-filtered, ¹³C-edited NOESY ($\tau_{\text{mix}} = 120$ ms) spectrum⁴⁴ of uniformly [¹⁵N, ¹³C]-labelled JARID1A-PHD3 bound to unlabelled H3_{1–9}K4me3 peptide was recorded. One-bond N-H RDCs were determined by using the in-phase anti-phase (IPAP) ¹⁵N-HSQC pulse sequence at 25 °C⁴⁵. Standard pulse sequences⁴⁶ were used for measurements of the ¹⁵N relaxation rates (R_1 , R_2) of JARID1A-PHD3 at 25 °C. The spectra were processed and analysed, respectively, with the NMRPipe⁴⁷ and Sparky (http://www.cgl.ucsf.edu/home/sparky) software. The solution structures of JARID1A-PHD3 both in the free state and in complex with H3_{1–9}K4me3 peptide were first calculated using the CYANA program⁴⁸. Interproton distance constraints were derived from 2D NOESY, 3D ¹⁵N-edited NOESY-HSQC and 3D ¹³C-edited NOESY-HSQC spectra. Backbone ϕ and ψ angles were derived from TALOS-based analysis of backbone chemical shifts⁴⁹. Several hydrogen bonds derived from chemical shift analysis and from observed NOEs characteristic for α -helices and β -sheets, were added in the final rounds of structure refinement. Of the 100 final structures calculated by CYANA, 20 structures with the lowest target functions were chosen for further refinement using the Xplor-NIH program⁵⁰, in which ¹D_{NH} RDC restraints, physical force field terms and explicit solvent terms⁵¹ were added to the calculation. The final structures were validated by Procheck-NMR⁵², and the statistics for the 20 final structures are listed in Supplementary Table 3.

Monomeric state of JARID1A-PHD3 in free and H3_{1–9}K4me3-bound states in solution. The oligomeric states of JARID1A-PHD3 (molecular weight 5.8 kDa) and JARID1A-PHD3–H3_{1–9}K4me3 complex (molecular weight 6.8 kDa) were first evaluated by comparing their elution volumes on a Superdex G75 16/60 column, with the calibration line derived from several molecular standards. As shown in Supplementary Fig. 7a, the elution volumes of both free JARID1A-PHD3 and the JARID1A-PHD3–H3_{1–9}K4me3 complex are comparable with those expected for their monomeric states, but considerably larger than those expected for their dimeric states. This suggests that JARID1A-PHD3 is monomeric in solution, for both the free and H3_{1–9}K4me3-bound states. Furthermore, the rotational correlation times of JARID1A-PHD3 in the free and H3_{1–9}K4me3-bound states were estimated as 3.5 ns and 4.5 ns, respectively, on the basis of an analysis of ¹⁵N R_2/R_1 relaxation time ratios (Supplementary Fig. 7b) using the quadratic diffusion program⁵³. These values are consistent with isotropic tumbling values of a monomeric protein of their respective sizes, providing further support that both free and complexed JARID1A-PHD3 exist as monomers in solution.

Thus, although the JARID1A-PHD3–H3K4me3 complex exhibits a domain-swapped dimer in the crystal, gel filtration and NMR relaxation measurements (Supplementary Fig. 7) clearly showed such a complex to be monomeric in solution. Hence the domain-swapped dimerization observed in the crystal is probably a characteristic feature of the crystalline state, originating perhaps in packing interactions.

31. Wang, G. G. *et al.* Quantitative production of macrophages or neutrophils *ex vivo* using conditional Hoxb8. *Nature Methods* **3**, 287–293 (2006).
32. Sawka-Verhelle, D. *et al.* PE-1/METS, an antiproliferative Ets repressor factor, is induced by CREB-1/CREM-1 during macrophage differentiation. *J. Biol. Chem.* **279**, 17772–17784 (2004).
33. Calvo, K. R., Sykes, D. B., Pasillas, M. & Kamps, M. P. Hoxa9 immortalizes a granulocyte-macrophage colony-stimulating factor-dependent promyelocyte capable of biphenotypic differentiation to neutrophils or macrophages, independent of enforced meis expression. *Mol. Cell. Biol.* **20**, 3274–3285 (2000).
34. Okada, Y. *et al.* hDOT1L links histone methylation to leukemogenesis. *Cell* **121**, 167–178 (2005).
35. Wallace, W., Schaefer, L. H. & Swedlow, J. R. A workingperson's guide to deconvolution in light microscopy. *Biotechniques* **31**, 1076–1078 (2001).
36. Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).
37. Fischle, W. *et al.* Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438**, 1116–1122 (2005).
38. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr.* **D 58**, 1948–1954 (2002).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr.* **D 60**, 2126–2132 (2004).
40. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* **D 53**, 240–255 (1997).
41. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr.* **D 63**, 32–41 (2007).
42. Brunger, A. T. Version 1.2 of the crystallography and NMR system. *Nature Protocols* **2**, 2728–2733 (2007).
43. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* **D 50**, 760–763 (1994).
44. Zwahlen, C. *et al.* Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N-peptide/boxB RNA complex. *J. Am. Chem. Soc.* **119**, 6711–6721 (1997).
45. Ottiger, M., Delaglio, F. & Bax, A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373–378 (1998).
46. Farrow, N. A. *et al.* Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994).
47. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
48. Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997).
49. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
50. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73 (2003).
51. Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M. & Nilges, M. Refinement of protein structures in explicit solvent. *Proteins* **50**, 496–506 (2003).
52. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
53. Lee, L. K., Rance, M., Chazin, W. J. & Palmer, A. G. III. Rotational diffusion anisotropy of proteins from simultaneous analysis of ¹⁵N and ¹³C alpha nuclear spin relaxation. *J. Biomol. NMR* **9**, 287–298 (1997).

LETTERS

Polar gradients of the DYRK-family kinase Pom1 couple cell length with the cell cycle

Sophie G. Martin¹ & Martine Berthelot-Grosjean¹

Cells normally grow to a certain size before they enter mitosis and divide. Entry into mitosis depends on the activity of Cdk1, which is inhibited by the Wee1 kinase and activated by the Cdc25 phosphatase¹. However, how cells sense their size for mitotic commitment remains unknown. Here we show that an intracellular gradient of the dual-specificity tyrosine-phosphorylation regulated kinase (DYRK) Pom1, which emanates from the ends of rod-shaped *Schizosaccharomyces pombe* cells, serves to measure cell length and control mitotic entry. Pom1 provides positional information both for polarized growth and to inhibit cell division at cell ends^{2–5}. We discovered that Pom1 is also a dose-dependent G2–M inhibitor. Genetic analyses indicate that Pom1 negatively regulates Cdr1 and Cdr2, two previously described Wee1 inhibitors of the SAD kinase family^{6–10}. This inhibition may be direct, because *in vivo* and *in vitro* evidence suggest that Pom1 phosphorylates Cdr2. Whereas Cdr1 and Cdr2 localize to a medial cortical region, Pom1 forms concentration gradients from cell tips that overlap with Cdr1 and Cdr2 in short cells, but not in long cells. Disturbing these Pom1 gradients leads to Cdr2 phosphorylation and imposes a G2 delay. In short cells, Pom1 prevents precocious M-phase entry, suggesting that the higher medial Pom1 levels inhibit Cdr2 and promote a G2 delay. Thus, gradients of Pom1 from cell ends provide a measure of cell length to regulate M-phase entry.

While studying *pom1* for its role in cell morphogenesis, we made several observations that indicate that *pom1* also regulates the cell cycle. First, *pom1Δ* cells divided at a significantly shorter size than wild-type cells (Table 1). Because fission yeast cells grow in length throughout interphase until mitosis, long size at division indicates mitotic delay, whereas short size indicates precocious mitosis. Second, *pom1* deletion partly suppressed the temperature-sensitivity of *cdc25-22* mutants, whereas moderate (less than fourfold) overexpression of Pom1 (*nmt1-pom1*) was synthetic sick with *cdc25-22* (Fig. 1b). Third, small increases in Pom1 levels caused a dose-dependent increase in cell size at division in both the wild-type and the *cdc25-22* mutants (Fig. 1c). Notably, even a <1.5-fold increase in Pom1 levels caused cells to divide at a significantly longer size. Thus, Pom1 is a dose-dependent inhibitor of G2–M transition.

To test whether this function is independent of the role of Pom1 in cell morphogenesis, we inhibited Pom1 in rod-shaped *cdc25-22* cells blocked in G2 by using a conditional *pom1-as1* allele that can be inhibited by the addition of the allele-specific inhibitor 1NM-PP1 (ref. 5). Fifty-two per cent of these cells were able to proceed through mitosis and septation at restrictive temperature, whereas most *cdc25-22 pom1⁺* cells remained blocked in G2 (Fig. 1d). Thus, Pom1 kinase activity is necessary to block G2–M transition independently of its role in cell morphogenesis.

To dissect how Pom1 inhibits the G2–M transition, we conducted epistasis experiments measuring the cell length at division of double mutants between *pom1Δ* and genes involved in G2–M control

(Table 1). Mutants in *cdc25* or members of the stress response pathway (*wis1* and *sty1*) that couples the cell cycle to nutrient availability^{11,12} were still responsive to altered Pom1 levels, suggesting that Pom1 does not regulate these pathways. In contrast, *wee1-50* mutants were not shortened by *pom1* deletion. In fact, *pom1Δ wee1-50* cells were slightly longer than *wee1-50* cells, possibly because of the extremely aberrant morphology of these double mutant cells (Supplementary Fig. 1). Wee1 is inhibited by a set of two related kinases: Cdr1 (also known as Nim1) and Cdr2 (refs 6–10). *cdr2Δ* was completely epistatic to *pom1Δ*, as the double mutant showed the same length at division as the single *cdr2Δ* mutant. Thus, Pom1 may negatively regulate Cdr2. Furthermore, *pom1Δ* was additive to *nif1Δ*, a

Table 1 | Genetic epistasis analysis indicates that Pom1 inhibits Cdr2

	YE	EMM	%	P value
Wild type (972)	12.7 ± 1.1 (312)	13.1 ± 1.3 (417)	100	—
<i>pom1Δ</i>	11.1 ± 2.0 (211)	11.7 ± 1.9 (580)	89	<10 ^{−10}
<i>pom1-2</i>	11.0 ± 1.6 (181)	11.0 ± 1.9 (227)	84	<10 ^{−14}
<i>pom1-GFP</i>	12.6 ± 0.9 (237)	13.0 ± 1.2 (264)	99	—
* <i>nmt1-pom1</i>	13.5 ± 1.5 (233)	15.7 ± 1.4 (131)	120	<10 ^{−26}
* <i>nmt1-pom1-GFP</i>	13.5 ± 1.4 (326)	16.4 ± 1.8 (480)	125	<10 ^{−59}
<i>cdc25-22</i>	19.2 ± 2.3 (116)	16.3 ± 1.6 (190)	124	<10 ^{−20}
<i>cdc25-22 pom1Δ</i>	14.6 ± 1.6 (129)	14.6 ± 1.6 (171)	111	—
<i>sty1Δ</i>	15.5 ± 1.4 (98)	—	—	<10 ^{−15}
<i>sty1Δ pom1Δ</i>	13.5 ± 1.7 (112)	—	—	—
<i>wis1Δ</i>	15.1 ± 2.3 (101)	—	—	<10 ^{−8}
<i>wis1Δ pom1Δ</i>	12.8 ± 2.6 (84)	—	—	—
<i>wis1DD</i>	11.2 ± 1.2 (109)	—	—	<10 ^{−29}
<i>wis1DD pom1Δ</i>	9.2 ± 1.1 (119)	—	—	—
<i>wee1-50</i>	6.6 ± 1.1 (414)	—	—	—
<i>wee1-50 pom1Δ</i>	7.7 ± 2.0 (504)	—	—	—
<i>cdr1Δ</i>	16.0 ± 1.1 (105)	18.2 ± 1.8 (138)	139	0.002
<i>cdr1Δ pom1Δ</i>	15.3 ± 1.5 (105)	17.4 ± 2.4 (119)	133	—
<i>cdr2Δ</i>	17.2 ± 1.2 (236)	20.1 ± 1.8 (233)	153	0.85
<i>cdr2Δ pom1Δ</i>	16.9 ± 2.6 (100)	20.1 ± 2.4 (190)	153	—
<i>cdr1Δ cdr2Δ</i>	15.8 ± 1.1 (101)	17.8 ± 1.3 (129)	134	0.93
<i>cdr1Δ cdr2Δ pom1Δ</i>	15.4 ± 1.4 (113)	17.8 ± 2.0 (103)	134	—
<i>nif1Δ</i>	11.7 ± 1.1 (557)	11.6 ± 1.2 (110)	89	<10 ^{−6}
<i>nif1Δ pom1Δ</i>	10.4 ± 2.3 (432)	10.6 ± 2.3 (151)	81	—
<i>pom1-2-GFP</i>	11.1 ± 1.6 (124)	11.8 ± 1.9 (105)	90	—
<i>pom1-cdr2C-GFP</i>	15.1 ± 0.9 (126)	16.0 ± 1.5 (136)	122	<10 ^{−31}
<i>pom1-cdr2C-GFP cdr2Δ</i>	16.8 ± 1.2 (114)	19.6 ± 1.2 (122)	150	—
<i>pom1-2-cdr2C-GFP</i>	11.4 ± 1.8 (219)	11.5 ± 1.6 (186)	88	—
<i>wee1-50 pom1-cdr2C-GFP</i>	5.9 ± 0.9 (224)	—	—	—
<i>pom1-tea1-GFP</i>	14.1 ± 0.9 (162)	15.7 ± 1.6 (141)	120	<10 ^{−25}
<i>pom1-tea1-GFP cdr2Δ</i>	15.3 ± 1.2 (138)	17.5 ± 1.7 (130)	134	—
<i>pom1-2-tea1-GFP</i>	11.0 ± 1.6 (217)	10.7 ± 1.4 (208)	82	—
<i>wee1-50 pom1-tea1-GFP</i>	6.3 ± 1.0 (146)	—	—	—

The average lengths of septated cells ± the s.d. within the cell population are shown. The numbers in parenthesis denote the number of cells measured. All strains were grown in complete (YE) medium at 25 °C, except for *wee1-50* strains, which were grown for 4 h at 36 °C. Prototrophic strains were also grown in minimal (EMM) medium at 25 °C. The length of prototrophic strains grown in EMM is shown as a percentage of wild-type. P values were determined by Student's *t*-test. Single mutants were compared to wild-type, whereas double and triple mutants were compared to appropriate *pom1⁺* single and double mutants. —, not determined.

* These strains were grown in EMM plus thiamine. All others were grown in EMM medium.

¹Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, Génopode Building, 1015 Lausanne, Switzerland.

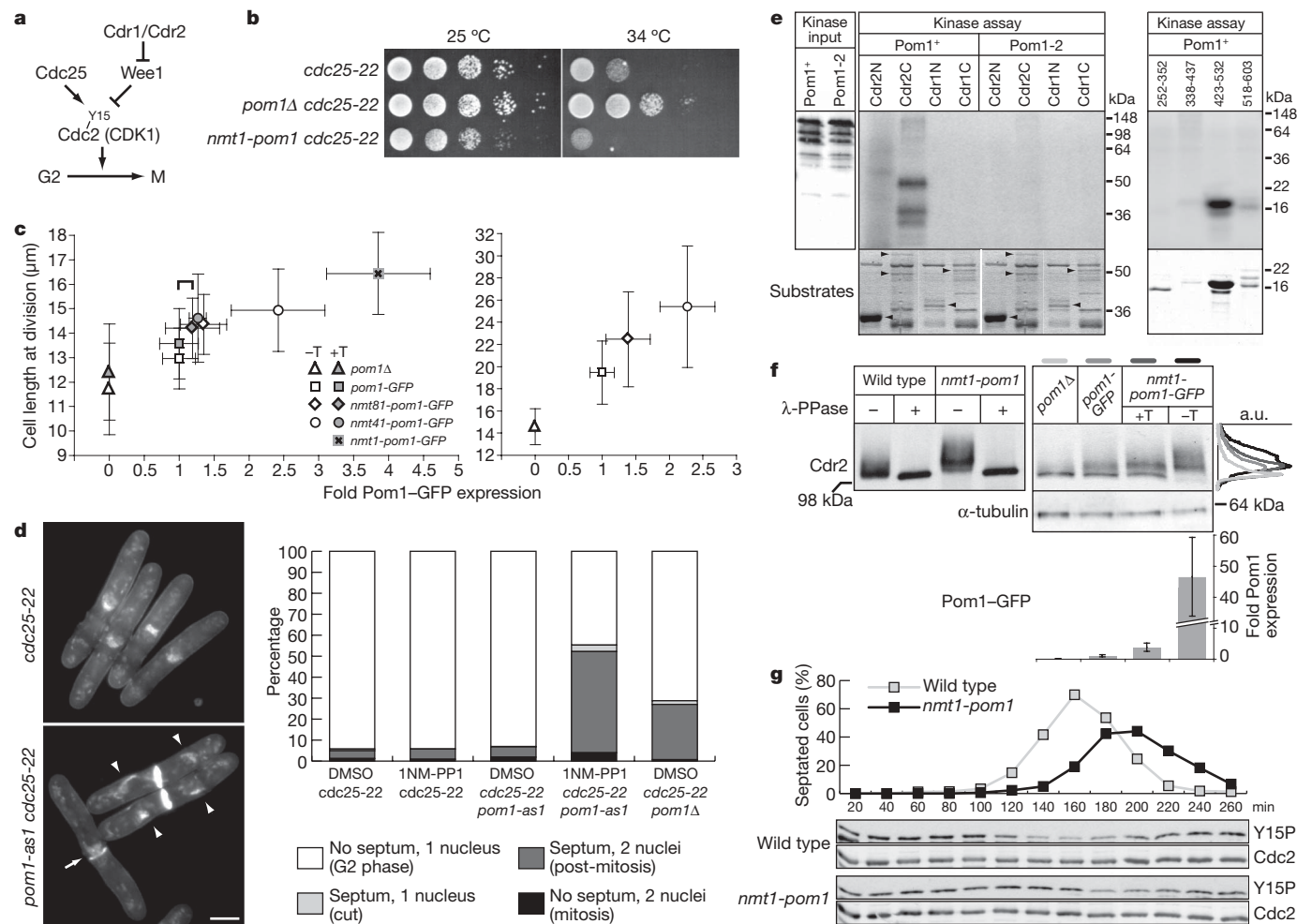


Figure 1 | Pom1 is a dose-dependent inhibitor of entry into mitosis and phosphorylates Cdr2. **a**, Regulation of G2–M transition. **b**, Serial dilutions of cells of indicated genotypes grown at indicated temperatures on complete medium (YE) plates, on which the *nmt1* promoter expresses only at basal levels. **c**, Dose-dependent effect of Pom1 on cell length at division. Indicated strains were grown in minimal medium (EMM) with or without thiamine (T), at 25°C, in wild-type (left) or *cdc25-22* (right) background. Analysis of variance (ANOVA) was highly significant for cell lengths and fluorescence intensities for both data sets ($P < 10^{-36}$ for all), and all pair-wise *t*-test comparisons within each growth condition were statistically significant ($P < 0.05$), except for the one noted by a bracket. There was a strong positive correlation between Pom1 levels and division length in both data sets ($r^2 = 0.88$, $P < 0.01$, and $r^2 = 0.98$, $P < 0.01$, respectively). **d**, Acute inhibition of Pom1 bypasses the G2 block of *cdc25-22* cells. Cells were incubated for 4 h at 36°C. After 3 h, 1NM-PP1 (or dimethylsulphoxide (DMSO) as control) was added to inhibit Pom1-as1. Nuclei and the cell wall were stained with 4,6-diamidino-2-phenylindole (DAPI) and calcofluor, respectively. Arrowheads indicate divided nuclei, the arrow points to 'cut' nucleus. Quantification is shown to the right ($n > 250$ for each genotype).

previously described inhibitor of Cdr1 (ref. 13), suggesting that Nif1 and Pom1 form two independent inhibitors of Cdr1 and Cdr2.

We tested whether Cdr1 or Cdr2 are Pom1 substrates. Pom1, but not kinase-dead Pom1 (ref. 14), efficiently phosphorylated fragments of the non-catalytic carboxy-terminal half of Cdr2 *in vitro* (Fig. 1e). In contrast, Pom1 failed to phosphorylate Cdr1 fragments. *In vivo*, western blot analysis showed slow-migrating species of Cdr2 that disappeared after phosphatase treatment, indicating Cdr2 phosphorylation. Phospho-Cdr2 levels were dependent on the dose of Pom1 (Fig. 1f). Thus, Cdr2 is phosphorylated in a *pom1*-dependent manner *in vivo*, and directly by Pom1 *in vitro*. Together with our genetic analyses, these data are consistent with a direct inhibition of Cdr2 function by Pom1.

Scale bar, 5 μm. **e**, *In vitro* kinase assays performed with immunoprecipitated Pom1–Myc or kinase-dead Pom1–2–Myc and recombinant Cdr1 and Cdr2 fragments. Anti-Myc western blot of kinase input (left), phosphorimager detection of 32 P incorporation (top) and Coomassie-stained gels of substrates input (bottom; same gel shown twice) are shown. Arrowheads point to specific bands. Sub-fragments of Cdr2C were used in the assay shown on the right. **f**, Cdr2 is phosphorylated in a *pom1*-dependent manner *in vivo*. The migration of Cdr2–2×HA–6×His with (+) or without (–) λ-phosphatase (λ-PPase) treatment (left) and in increasing concentrations of Pom1–GFP (right) was detected by anti-haemagglutinin (HA) antibodies. The Cdr2 intensity profiles along each lane are shown on the far right: higher Pom1 dose is indicated by a darker shade of grey. α-tubulin was used as a loading control. The graph (bottom) shows Pom1–GFP levels as measured by GFP fluorescence intensity. a.u., arbitrary units. **g**, Elutriation of wild-type and *nmt1-pom1* strains grown in EMM plus thiamine. The septation index was measured on calcofluor-stained cells. Blots show the levels of Cdc2 Tyr 15 phosphorylation (Y15P) and total Cdc2 as loading control. All error bars are s.d.

The endpoint of the Cdr1/Cdr2–Wee1 pathway is inhibition of Cdc2 by phosphorylation at Tyr 15 (refs 15, 16) (*cdc2* is the *S. pombe* CDK1 gene; Fig. 1a). If Pom1 negatively regulates Cdr2, this predicts that Pom1 overexpression will lead to prolonged Wee1 activation and delay Cdc2 dephosphorylation. To test this hypothesis, we synchronized wild-type and *nmt1-pom1* cells by elutriation and assayed the levels of Cdc2 Tyr 15 phosphorylation over time (Fig. 1g). Compared to wild-type, *nmt1-pom1* cells showed a 40 min delay in Cdc2 dephosphorylation with a correlating delay in septation. Thus, the delay in mitosis entry observed in *pom1*-overexpressing cells is probably due to delayed Cdc2 activation.

We examined the localization of Pom1, Nif1, Cdr1 and Cdr2 by tagging each gene at its endogenous chromosomal locus. Pom1

localizes to cell tips^{2,14}. We found that Nif1 also localized to cell tips, although through distinct mechanisms, and was often enriched at the opposite tip from Pom1 (Supplementary Fig. 2). In contrast, Cdr1 and Cdr2 formed a medial broad band of cortical dots throughout interphase (Supplementary Fig. 3)¹⁷. Cdr1 and Cdr2 became diffuse during mitosis. Cdr2 also transiently localized to the cytokinetic ring. Cdr1 localization depended on Cdr2, whereas Cdr2 localization was independent of Cdr1 (Supplementary Fig. 4). This places Cdr2 upstream of Cdr1, at least in terms of localization. Thus, Cdr2 and Pom1 occupy distinct cortical domains (Supplementary Fig. 4).

We investigated the interdependence between Cdr1/Cdr2 and Pom1/Nif1 localization. Pom1 and Nif1 localized normally to cell tips in *cdr1Δ* and *cdr2Δ* cells (Supplementary Fig. 4). In contrast, Cdr1 and Cdr2 spread over the entire non-growing half of *pom1Δ* cells (Fig. 2a and data not shown). Mislocalization of Cdr2 did not correlate directly with aberrant growth patterns: whereas Cdr2 was similarly mislocalized in *tea4Δ* cells and, to a smaller extent, in *tea1Δ* cells, in which Pom1 fails to localize to cell tips^{2,5,18–20}, it localized

correctly at the cell middle in other monopolar mutants, such as *tea3Δ*, *bud6Δ* or *orb2-34* (Fig. 2b and Supplementary Fig. 5). This localization is reminiscent of that of Mid1 (refs 3, 5), and indicates the existence of complementary mechanisms to restrict Cdr2 from the growing cell end. We note that Nif1 is dispensable for Cdr1 and Cdr2 localization in wild-type or *pom1Δ* background (data not shown). Thus, Pom1 inhibits Cdr2 localization at the non-growing cell end.

Although *tea1Δ*, *tea4Δ* and *pom1Δ* mutants showed the same Cdr2 expansion at the non-growing cell end (Fig. 2a, b), *tea1Δ* and *tea4Δ* mutants did not cause early mitosis like *pom1Δ*. In contrast, *tea1Δ* and *tea4Δ* were synthetic sick with *cdc25-22*, similar to Pom1-over-expressing cells, suggesting a modest G2 delay (Fig. 2d). This synthetic sickness was due, at least in part, to the presence of Pom1, as deletion of *pom1* improved the growth defect of *tea1Δ cdc25-22* and *tea4Δ cdc25-22* double mutants. Thus, Pom1 is active in *tea1Δ* and *tea4Δ* mutants¹⁴ and functions to inhibit cell cycle progression. This suggests that delocalized Pom1 in *tea1Δ* and *tea4Δ* mutants may cause Cdr2 inhibition and G2 delay.

To test this hypothesis further, we targeted Pom1 ectopically to the cell middle by tagging it at its endogenous chromosomal locus with protein or protein fragment–GFP (green fluorescent protein) fusions. We first used the non-catalytic C-terminal region of Cdr2 (Cdr2C) to drive Pom1 to the cell middle. Pom1–Cdr2C–GFP could be observed at cell ends, similar to wild-type Pom1, and weakly at medial cortical dots reminiscent of the localization of Cdr2 (Fig. 2c). This fusion did not cause important morphological abnormalities (Supplementary Fig. 1). However, *pom1-cdr2C-GFP* cells were significantly longer than *pom1-GFP* cells and synthetic sick with *cdc25-22*, indicating a G2 delay (Table 1 and Fig. 2d). This phenotype was dependent on active ectopic Pom1, as it was completely reversed by abolishing Pom1 kinase activity with a point mutation in its catalytic domain (Table 1 and Fig. 2d). Notably, both fusions were expressed

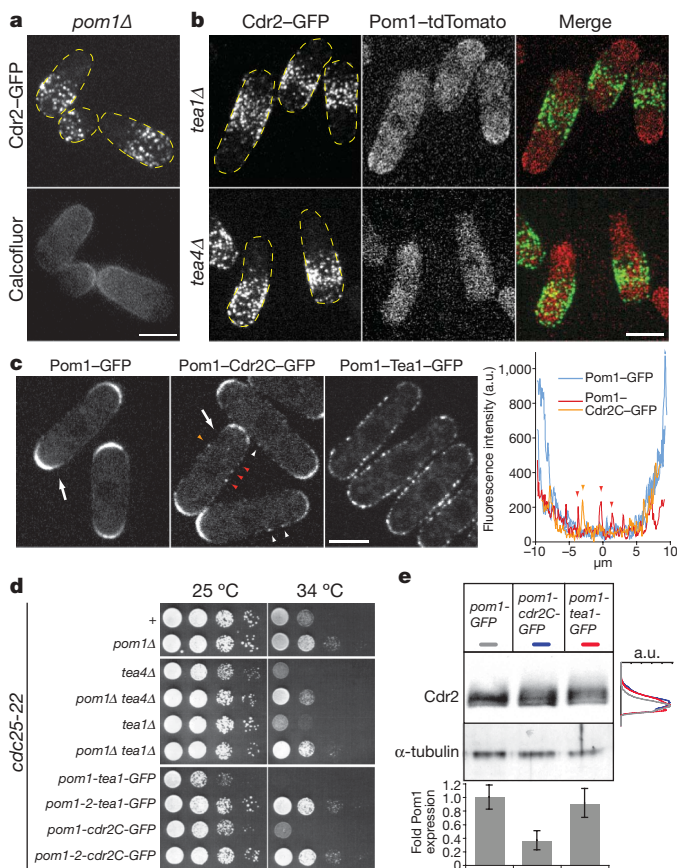


Figure 2 | Ectopic localization of Pom1 to the cell middle imposes a G2 delay. **a**, Maximum projection of spinning disk confocal sections of Cdr2–GFP and wide-field calcofluor image in *pom1Δ* cells. **b**, Maximum projection of spinning disk confocal sections of Pom1–tdTomato and Cdr2–GFP in *tea1Δ* and *tea4Δ* cells. Contrasts were increased to detect delocalized Pom1–tdTomato. **c**, Single confocal section at the cell equator of Pom1–GFP, Pom1–Cdr2C–GFP and Pom1–Tea1–GFP. The intensity profiles of Pom1–GFP and Pom1–Cdr2C–GFP along the cortex of the cells marked with an arrow are shown on the right (each cell side shown as an independent trace). Ectopic peaks of higher intensity at the cell middle (arrowheads) are detected in Pom1–Cdr2C–GFP cells. All scale bars, 5 μm. **d**, Serial dilutions of *cdc25-22* cells of indicated genotype were grown at the permissive (25 °C) and semi-restrictive (34 °C) temperature for *cdc25-22*. **e**, Western blot analysis of Cdr2–2×HA–6×His in indicated genotypes with anti-HA antibodies and α-tubulin as a loading control. The intensity profile of the Cdr2 blot along each lane is shown on the right. The graph (bottom) shows the expression levels of Pom1 fusions measured by GFP fluorescence intensity. Error bars denote s.d.

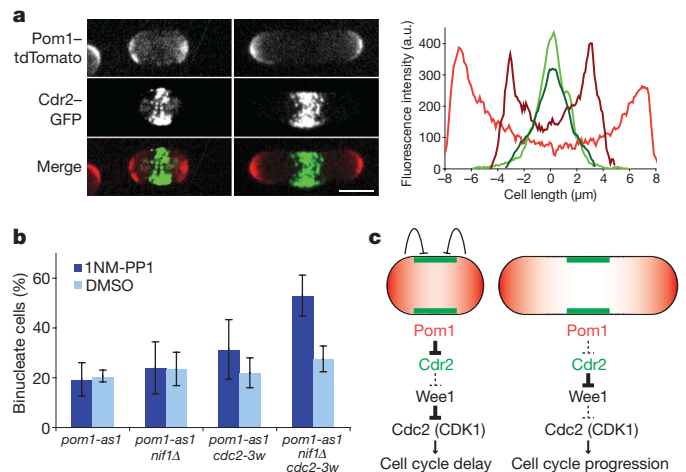


Figure 3 | Pom1 overlaps with Cdr2 and prevents mitosis in small cells. **a**, Sum projection of spinning disk confocal sections of Pom1–tdTomato and Cdr2–GFP localization in short and long wild-type cells. The average fluorescence intensity of Pom1–tdTomato (red) and Cdr2–GFP (green) over the length of three cells of identical length is shown to the right. Lighter colours are used for the longer cells. Measurements on 20 other individual cells are shown in Supplementary Fig. 8. Scale bar, 5 μm. **b**, Pom1 prevents precocious mitosis in small cells. The percentage of binucleate (septated or non-septated) cells was quantified after 1 h inhibition of Pom1–as1 with 1NM-PP1 at 30 °C in cells of indicated genotype. DMSO was used as control. Error bars denote s.d. **c**, Model for how fission yeast cells monitor their length to control entry into mitosis. A gradient of Pom1 from cell ends overlaps significantly with Cdr2 at the middle of small cells, leading to Cdr2 inhibition and cell cycle delay. Once the cells reach a longer size, the concentration of Pom1 at the cell middle is no longer sufficient for significant Cdr2 inhibition, allowing cell cycle progression.

at similar low levels, and the kinase-dead Pom1 fusion was also targeted to the cell middle (Fig. 2e and Supplementary Fig. 6). The increased length of *pom1-cdr2C-GFP* cells was completely dependent on *wee1* and *cdr2*, suggesting that ectopic Pom1 regulates the *wee1* pathway through *cdr2* (Table 1). In agreement with this idea, we also observed a modest but reproducible increase in the level of phosphorylated Cdr2 on western blot (Fig. 2e and Supplementary Fig. 7). Thus, ectopic localization of active Pom1 to the cell middle may impose a G2 delay by phosphorylating and inactivating Cdr2.

An independent *pom1-tea1-GFP* fusion was constructed, initially to retarget Pom1 to the cell end. However, for unknown reasons this fusion protein localized in a patchy pattern all over the cell cortex (Fig. 2c). As with the *cdr2C* fusion, *pom1-tea1-GFP* cells showed a Pom1 activity-dependent increase in cell length and synthetic sickness with *cdc25-22*, and showed modest increase in the levels of phosphorylated Cdr2 (Table 1, Fig. 2d, e and Supplementary Fig. 7). Epistasis experiments showed that the length increase caused by *pom1-tea1-GFP* was not additive to the *cdr2Δ* phenotype, suggesting that delocalized Pom1 functions through Cdr2. However, because *cdr2Δ pom1-tea1-GFP* double mutants were shorter than *cdr2Δ* single mutants, this fusion may have further Cdr2-independent effects on cell growth. Accordingly, the *pom1-tea1-GFP* fusion also caused some morphological abnormalities (Supplementary Fig. 1).

These experiments suggest that an overlap between Pom1 and Cdr2 leads to Cdr2 inhibition. We tested whether such overlap occurs in wild-type cells. Pom1 forms gradients from cell ends⁵. Wild-type cells are born by medial fission at 7–8 μm in length and grow by tip extension throughout the cell cycle until they reach about 14 μm and divide. Pom1 gradients in short (<8 μm) and long (>13 μm) cells were of similar shape such that the amount of Pom1 in the cell middle was significantly higher in small compared to large cells (Fig. 3a and Supplementary Fig. 8). In contrast, Cdr2 distribution was not significantly altered in short versus long cells. Thus, there is a higher overlap of Pom1 with Cdr2 in short compared to long cells.

This localization data suggested that Pom1 may be particularly important in small cells. We therefore tested *pom1* function in small cells of the *cdc2-3w* allele, which no longer require *cdc25⁺* for mitotic induction but strongly depend on *wee1⁺* function²¹. We also removed the function of the second Cdr1/Cdr2 inhibitor Nif1 (ref. 13). To avoid indirect effects due to the *pom1* function in cell morphogenesis, we inhibited Pom1 activity using the *pom1-as1* allele for a short time period. Notably, whereas inactivation of Pom1 for 1 h had no significant effect in wild-type or *nif1Δ* mutant cells, more than 50% of *cdc2-3w nif1Δ* cells became binucleate, indicating that Pom1 inhibition provokes precocious mitotic entry (Fig. 3b). These results thus indicate that Pom1 is important to delay entry into M phase in short cells that lack complementary mechanisms of length control.

We propose that Pom1 is a new size sensor that couples cell length with cell cycle progression. Pom1 forms concentration gradients from cell ends. In short cells, we suggest that Pom1 levels in the cell middle cause Cdr2 inhibition and cell cycle delay. As cells grow, Pom1 levels at the cell middle diminish. Once the cell has reached a critical length, medial Pom1 levels no longer inhibit Cdr2 significantly, thus allowing cell cycle progression (Fig. 3c). Notably, less than twofold changes in medial Pom1 levels (Fig. 1c and Supplementary Fig. 9) are sufficient to delay the cell cycle, suggesting that Pom1 levels and gradient shape are tightly regulated. As *pom1* is non-essential in cells with otherwise normal cell cycle regulation, this new mechanism may act additively with other cell growth monitoring systems, including translational control of G1–S and G2–M cell cycle regulators²².

Gradients are used to regulate several intracellular processes. During cell cycle progression, a Ran-GTP gradient around chromosomes organizes the mitotic spindle²³. At the spindle midzone, Aurora B kinase then generates a gradient of post-translational marks that provides spatial information for anaphase and cytokinesis²⁴. Pom1 gradients, which may be shaped by coupled localization to cell

ends and diffusion, provide an appealing device for a cell to continuously monitor its length by responding to concentration changes at the cell middle. Many DYRKs function to inhibit the cell cycle and promote differentiation^{25–27}. SAD kinases are essential for neuronal polarization²⁸ and function in the budding yeast to couple Wee1 with cell morphogenesis²⁹. A future challenge will be to determine whether a similar intracellular gradient system also functions in multicellular organisms to link proliferation with cell size.

METHODS SUMMARY

Strains used in this study are listed in Supplementary Table 1. Standard methods for microscopy, molecular biology and biochemistry were used throughout and are described in the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 January; accepted 9 April 2009.

Published online 27 May 2009.

1. Nurse, P. Universal control mechanism regulating onset of M-phase. *Nature* **344**, 503–508 (1990).
2. Böhler, J. & Pringle, J. R. Pom1p, a fission yeast protein kinase that provides positional information for both polarized growth and cytokinesis. *Genes Dev.* **12**, 1356–1370 (1998).
3. Celton-Morizur, S., Racine, V., Sibarita, J. B. & Paoletti, A. Pom1 kinase links division plane position to cell polarity by regulating Mid1p cortical distribution. *J. Cell Sci.* **119**, 4710–4718 (2006).
4. Huang, Y., Chew, T. G., Ge, W. & Balasubramanian, M. K. Polarity determinants Tea1p, Tea4p, and Pom1p inhibit division-septum assembly at cell ends in fission yeast. *Dev. Cell* **12**, 987–996 (2007).
5. Padte, N. N., Martin, S. G., Howard, M. & Chang, F. The cell-end factor Pom1p inhibits Mid1p in specification of the cell division plane in fission yeast. *Curr. Biol.* **16**, 2480–2487 (2006).
6. Coleman, T. R., Tang, Z. & Dunphy, W. G. Negative regulation of the *wee1* protein kinase by direct action of the *nim1/cdr1* mitotic inducer. *Cell* **72**, 919–929 (1993).
7. Parker, L. L., Walter, S. A., Young, P. G. & Piwnicka-Worms, H. Phosphorylation and inactivation of the mitotic inhibitor Wee1 by the *nim1/cdr1* kinase. *Nature* **363**, 736–738 (1993).
8. Wu, L. & Russell, P. Nim1 kinase promotes mitosis by inactivating Wee1 tyrosine kinase. *Nature* **363**, 738–741 (1993).
9. Breeding, C. S. et al. The *cdc2⁺* gene encodes a regulator of G₂/M progression and cytokinesis in *Schizosaccharomyces pombe*. *Mol. Biol. Cell* **9**, 3399–3415 (1998).
10. Kanoh, J. & Russell, P. The protein kinase Cdr2, related to Nim1/Cdr1 mitotic inducer, regulates the onset of mitosis in fission yeast. *Mol. Biol. Cell* **9**, 3321–3334 (1998).
11. Petersen, J. & Hagan, I. M. Polo kinase links the stress pathway to cell cycle control and tip growth in fission yeast. *Nature* **435**, 507–512 (2005).
12. Shiozaki, K. & Russell, P. Cell-cycle control linked to extracellular environment by MAP kinase pathway in fission yeast. *Nature* **378**, 739–743 (1995).
13. Wu, L. & Russell, P. Nif1, a novel mitotic inhibitor in *Schizosaccharomyces pombe*. *EMBO J.* **16**, 1342–1350 (1997).
14. Böhler, J. & Nurse, P. Fission yeast Pom1p kinase activity is cell cycle regulated and essential for cellular symmetry during growth and division. *EMBO J.* **20**, 1064–1073 (2001).
15. Lundgren, K. et al. mik1 and wee1 cooperate in the inhibitory tyrosine phosphorylation of cdc2. *Cell* **64**, 1111–1122 (1991).
16. Parker, L. L., Atherton-Fessler, S. & Piwnicka-Worms, H. p107^{wee1} is a dual-specificity kinase that phosphorylates p34^{cdc2} on tyrosine 15. *Proc. Natl Acad. Sci. USA* **89**, 2917–2921 (1992).
17. Morrell, J. L., Nichols, C. B. & Gould, K. L. The GIN4 family kinase, Cdr2p, acts independently of septins in fission yeast. *J. Cell Sci.* **117**, 5293–5302 (2004).
18. Tatebe, H., Shimada, K., Uzawa, S., Morigasaki, S. & Shiozaki, K. Wsh3/Tea4 is a novel cell-end factor essential for bipolar distribution of Tea1 and protects cell polarity under environmental stress in *S. pombe*. *Curr. Biol.* **15**, 1006–1015 (2005).
19. Martin, S. G., McDonald, W. H., Yates, J. R. III & Chang, F. Tea4p links microtubule plus ends with the formin for3p in the establishment of cell polarity. *Dev. Cell* **8**, 479–491 (2005).
20. Mata, J. & Nurse, P. tea1 and the microtubular cytoskeleton are important for generating global spatial order within the fission yeast cell. *Cell* **89**, 939–949 (1997).
21. Russell, P. & Nurse, P. Negative regulation of mitosis by *wee1⁺*, a gene encoding a protein kinase homolog. *Cell* **49**, 559–567 (1987).
22. Tapon, N., Moberg, K. H. & Hariharan, I. K. The coupling of cell growth to the cell cycle. *Curr. Opin. Cell Biol.* **13**, 731–737 (2001).
23. Kalab, P., Pralle, A., Isacoff, E. Y., Heald, R. & Weis, K. Analysis of a RanGTP-regulated gradient in mitotic somatic cells. *Nature* **440**, 697–701 (2006).
24. Fuller, B. G. et al. Midzone activation of aurora B in anaphase produces an intracellular phosphorylation gradient. *Nature* **453**, 1132–1136 (2008).

25. Souza, G. M., Lu, S. & Kuspa, A. YAKA, a protein kinase required for the transition from growth to development in *Dictyostelium*. *Development* **125**, 2291–2302 (1998).
26. Pellettieri, J., Reinke, V., Kim, S. K. & Seydoux, G. Coordinate activation of maternal protein degradation during the egg-to-embryo transition in *C. elegans*. *Dev. Cell* **5**, 451–462 (2003).
27. Mercer, S. E. & Friedman, E. Mirk/Dyrk1B: a multifunctional dual-specificity kinase involved in growth arrest, differentiation, and cell survival. *Cell Biochem. Biophys.* **45**, 303–315 (2006).
28. Kishi, M., Pan, Y. A., Crump, J. G. & Sanes, J. R. Mammalian SAD kinases are required for neuronal polarization. *Science* **307**, 929–932 (2005).
29. Lew, D. J. The morphogenesis checkpoint: how yeast cells watch their figures. *Curr. Opin. Cell Biol.* **15**, 648–653 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We wish to thank J. Bähler, F. Chang, P. Cruz, K. Gull, I. Hagan, P. Russell, K. Shiozaki and V. Simanis for strains and reagents. We thank R. Benton, F. Chang, C. Fankhauser, N. Hernandez and F. Bendezu for critical reading of the manuscript. We also wish to thank J. Moseley and P. Nurse for open discussion before publication. Research in the laboratory of S.G.M. is supported by a Swiss National Science Foundation Professorship grant (PP00A–114936) and a Human Frontiers Science Program Career Development Award (CDA0016/2008).

Author Contributions S.G.M. conceived the project, and designed and analysed the experiments. S.G.M. and M.B.G. performed the experiments. The manuscript was written by S.G.M. with input from M.B.G.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.G.M. (Sophie.Martin@unil.ch).

METHODS

Yeast strains, media and genetic methods. Standard methods for *S. pombe* media and genetic manipulations were used throughout. Tagged, deletion and promoter swap strains were constructed by using a PCR-based approach³⁰ and confirmed by PCR. Templates for tdTomato and 3×GFP tagging were used as published^{31,32}. *pom1* under control of the *nmt* promoter was either grown with 5 µg ml⁻¹ thiamine (Sigma) or induced to steady-state levels for >24 h in medium lacking thiamine, except for Fig. 1f in which cells were induced for 20 h.

Pharmacological inhibitors and dyes. 1NM-PP1 (Calbiochem) was used at a final concentration of 10 µM from a 10 mM stock solution in DMSO, and added for 1 h to cultures in YE medium, as we found that 1NM-PP1 was more efficient in YE than EMM medium. Methyl-2-benzimidazolecarbamate (Aldrich) and Latrunculin A (kindly provided by P. Cruz, University of California) were added to cultures for 20 min at a final concentration of 25 µg ml⁻¹ and 200 µM, respectively, from 100× stock solutions in DMSO. Calcofluor (Sigma) was added at a final concentration of 5 µg ml⁻¹ from a 200× stock solution. DAPI was added at a final concentration of 1 µg ml⁻¹ to ethanol-fixed cells.

Microscopy and image analysis. Microscopy was performed at room temperature (23 °C) with either a spinning-disk or a wide-field fluorescence microscope. Spinning disk microscopy was carried out using a Leica DM14000B inverted microscope equipped with an HCX PL APO ×100/1.46 numerical aperture (NA) oil objective and a PerkinElmer Ultraview Confocal system (including a Yokagawa CSU22 real-time confocal scanning head, an argon/krypton laser and a cooled 14-bit frame transfer EMCCD C9100-50 camera). Stacks of z-series confocal sections were acquired at 0.3-µm intervals with the Ultraview software, and images were rendered by two-dimensional maximum-intensity projection unless otherwise indicated. Wide-field microscopy was performed on a Leica AF6000 system consisting of a DM6000B upright microscope fitted with an HCX PL APO ×63/1.3 NA glycerol objective, a Leica DFC350x CCD camera, a Leica EL6000 light source, and Chroma filter sets. Images were acquired with the Leica AS AF software. Figures were prepared with Adobe Photoshop CS3 and Adobe Illustrator CS3.

For cell-length measurements, all strains were grown in YE or EMM medium at 25 °C, except for *wee1-50* strains, which were shifted to 36 °C for 4 h. For imaging of tagged proteins, strains were grown in EMM with appropriate supplements.

Fluorescence quantification. For whole-cell quantification of the distribution of tagged Pom1 and Cdr2, a sum projection of spinning disk confocal z-stacks of an individual cell was boxed and the ImageJ (1.30v) 'plot profile' tool was used to compress the fluorescence intensity into a one-dimensional line along the long axis of the cell, and generate a fluorescence profile that measures the total fluorescence intensity for each point along the cell length. Background correction was performed by subtracting the background fluorescence intensity measured along one identical box containing no cell on either side of the cell examined. We did not correct for autofluorescence, because the autofluorescence is not uniform along the cell length, and the autofluorescence of wild-type cells expressing no fluorescent marker was very low under the imaging conditions used. Thus, the quantifications shown in Fig. 3a and Supplementary Figs 8 and 9 probably slightly underestimate the changes in Pom1-GFP intensity at the cell middle. To confirm the distribution of Pom1 by independent imaging technique, we also quantified the fluorescence intensity along the length of cells imaged in a single epifluorescence focal plane (Supplementary Fig. 8). For quantification of cortical levels of Pom1 (Fig. 2c), we measured the intensity profile along a cortical line drawn around the cell cortex and subtracted the background intensity.

For measurement of global Pom1 levels in wild-type, overexpression and fusion strains, we measured the average fluorescence intensity of sum projections of spinning disk confocal z-stacks of at least eight individual cells for each genotype. Background correction was performed by subtracting the background fluorescence intensity in a region that did not contain cells, and the autofluorescence levels of wild-type cells expressing no fluorescent marker imaged and processed in the same conditions. All values were normalized to that of wild-type Pom1-GFP.

Construction of Pom1 fusions. All plasmids were constructed using standard molecular biology techniques. In general, genes or gene fragments were cloned after PCR from genomic DNA with primers containing 5' extensions with specific restriction sites. Details of the primers and restriction sites used are available on request. pSM752 contains in this order 500 base pairs (bp) of *pom1* 3'

untranslated region (UTR), a unique NotI restriction site, the final 500 bp of the *pom1* open reading frame (ORF; without stop codon), 1.6 kb of *cdr2* encoding amino acids 253–776, GFP and a *ura4⁺* selective marker. The *pom1* ORF, *cdr2C* and GFP were all in frame. The plasmid was linearized with NotI, transformed into YSM1183 and integrated at the *pom1* locus. A similar procedure yielded plasmid pSM735 for tagging *pom1* with *tea1-GFP*. This procedure results in tagging a single copy of *pom1* under control of the endogenous promoter. For tagging of kinase-dead *pom1*, the same constructs were integrated into a *pom1-2* mutant.

In vitro kinase assay. Cdr1 and Cdr2 fragments were cloned in pRSETA (Invitrogen), expressed in BL21 cells and purified with HisPure (Pierce) according to the manufacturer's protocol. Cdr2 fragments were: 1–269 (Cdr2N), 253–775 (Cdr2C), 252–352, 338–437, 423–532 and 518–603. Smaller fragments at the far C terminus were not expressed. Cdr1 fragments were: Cdr1N (amino acids 1–270) and Cdr1C (amino acids 234–594). Owing to poor solubility of the fragments expressed, only low amounts of proteins were recovered. Extracts from yeast grown in EMM medium were prepared in CXS buffer (50 mM HEPES, pH 7.0, 20 mM KCl, 1 mM MgCl₂, 2 mM EDTA, pH 7.5, and protease inhibitor cocktail) by grinding in liquid nitrogen with a mortar and pestle as described¹⁹. After thawing, Triton X-100 was added to a final concentration of 0.1%. For immunoprecipitations, 150 µl soluble extract was added to 20 µl sheep anti-mouse magnetic Dynabead slurry (Dyna) pre-bound (non-covalently) to 2 µg monoclonal anti-Myc antibodies (9E10; Santa Cruz Biotech), and incubated for 2 h at 4 °C. After incubation, the beads were washed twice with CXS 350 mM NaCl, four times with CXS 150 mM NaCl, and once with kinase buffer (50 mM HEPES, pH 7.4, 1 mM EDTA, 5% glycerol, 150 mM NaCl, 10 mM MgCl₂ and 10 mM MnCl₂). Kinase buffer and substrate were mixed directly with the beads and the reaction was started by addition of [γ -³²P]ATP (PerkinElmer). After a 20-min incubation at 30 °C, the reaction was stopped by boiling in sample buffer and analysed by SDS-PAGE. ³²P-incorporation was detected in a phosphorimager.

Elutriation. Wild-type (972) and *nmt1-pom1* (YSM1318) strains were grown to exponential phase in EMM plus thiamine at 25 °C. Elutriation was performed using a Beckman JE5.0 rotor with a 40-ml elutriation chamber. Cell cultures were loaded at 728g and 50 ml min⁻¹ and an 800-ml fraction was eluted with constant flow by lowering the rotation speed to 702g for wild-type and 488g for *nmt1 pom1*. This fraction was returned to 25 °C and aliquots were collected every 20 min for calcofluor staining and protein extraction. For extract preparation, cells were washed once in stop buffer (1 mM NaN₃, 10 mM EDTA, 50 mM NaF, 150 mM NaCl), pelleted and stored at –80 °C. Cells were broken with acid-washed glass beads (Sigma) in CXS buffer supplemented with 0.1% NP-40 and phosphatase inhibitor cocktail (Sigma). Total protein concentration was measured with a BCA assay (Pierce) and equal amounts were loaded on SDS-PAGE.

Western blot analysis. For analysis of Cdr2 phosphorylation *in vivo*, extracts were prepared by grinding in liquid nitrogen using 6×His denaturing buffer (50 mM sodium phosphate, 300 mM NaCl, 6 M guanidine, 10 mM imidazole, pH 7.4). Protein concentration was measured with a BCA assay (Pierce) and by western blotting with an anti- α -tubulin antibody. Equal amounts of proteins were mixed with HisPure resin (Pierce) and incubated for 30 min at room temperature. The resin was washed once with 6×His non-denaturing buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.4), resuspended in sample buffer and analysed using standard SDS-PAGE and western blotting procedures. For λ -phosphatase treatment, the Cdr2-2×HA-6×His-bound resin was incubated for 1 h at 30 °C in the presence or absence of λ -phosphatase (NEB). Antibodies used on western blots were: mouse monoclonal anti-HA (HA.11; Covance), anti-GFP (Roche), anti-Myc (9E10; Santa Cruz Biotech), anti-His (Qiagen), anti-Cdc2 (PSTAIR) (sc-53; Santa Cruz Biotechnology), anti-phospho-Cdc2 (Tyr15) (9111; Cell Signaling) and anti- α -tubulin (TAT1, a kind gift from K. Gull).

30. Bähler, J. *et al.* Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14**, 943–951 (1998).
31. Snaith, H. A., Samejima, I. & Sawin, K. E. Multistep and multimode cortical anchoring of tea1p at cell tips in fission yeast. *EMBO J.* **24**, 3690–3699 (2005).
32. Wu, J. Q. *et al.* Assembly of the cytoskeletal contractile ring from a broad band of nodes in fission yeast. *J. Cell Biol.* **174**, 391–402 (2006).

A spatial gradient coordinates cell size and mitotic entry in fission yeast

James B. Moseley¹, Adeline Mayeux², Anne Paoletti² & Paul Nurse¹

Many eukaryotic cell types undergo size-dependent cell cycle transitions controlled by the ubiquitous cyclin-dependent kinase Cdk1 (refs 1–4). The proteins that control Cdk1 activity are well described but their links with mechanisms monitoring cell size remain elusive. In the fission yeast *Schizosaccharomyces pombe*, cells enter mitosis and divide at a defined and reproducible size owing to the regulated activity of Cdk1 (refs 2, 3). Here we show that the cell polarity protein kinase Pom1, which localizes to cell ends⁵, regulates a signalling network that contributes to the control of mitotic entry. This network is located at cortical nodes in the middle of interphase cells, and these nodes contain the Cdk1 inhibitor Wee1, the Wee1-inhibitory kinases Cdr1 (also known as Nim1) and Cdr2, and the anillin-like protein Mid1. Cdr2 establishes the hierarchical localization of other proteins in the nodes, and receives negative regulatory signals from Pom1. Pom1 forms a polar gradient extending from the cell ends towards the cell middle and acts as a dose-dependent inhibitor of mitotic entry, working through the Cdr2 pathway. As cells elongate, Pom1 levels decrease at the cell middle, leading to mitotic entry. We propose that the Pom1 polar gradient and the medial cortical nodes generate information about cell size and coordinate this with mitotic entry by regulating Cdk1 through Pom1, Cdr2, Cdr1 and Wee1.

Recurring cycles of cellular growth and division present a simple question: how do cells link these processes and divide only upon reaching an appropriate size? Genetic screens have identified components in the fission yeast controlling cell size at division^{6,7}. The central factor Cdk1 drives entry into mitosis by phosphorylating a wide range of targets⁸. In small cells that are in early G2, Cdk1 is kept inactive by the kinase Wee1, which directly phosphorylates and inhibits Cdk1. As cell size increases during G2, the phosphatase Cdc25 removes this inhibitory phosphorylation to activate Cdk1 and allow mitotic entry. Thus, the mitotic entry control system coordinates a balance of Wee1 and Cdc25 activity with changes in cell size to generate a reproducible cell size at division. Although several factors regulating Wee1 and Cdc25 have been identified³, the links between cell size and the Wee1–Cdc25–Cdk1 module have remained elusive.

We noted that the protein kinase Cdr2, which negatively regulates Wee1, is localized to a band of cortical nodes in the middle of interphase cells⁹. After mitotic entry, cytokinesis factors such as myosin II are recruited to similar nodes, which subsequently condense to form the cytokinetic ring¹⁰. The presence of Cdr2 in these medial nodes during G2, when the inhibition of Wee1 contributes to the G2–M transition, suggested that the nodes might act as potential sites of cell cycle regulation. To investigate this possibility further, we first analysed the composition of medial nodes. We found that the previously uncharacterized protein Blt1 (also known as SPBC1A4.05), identified using a proteomic approach (Supplementary Fig. 1), co-localized with Cdr2 in the medial interphase nodes, together with Mid1 (Fig. 1a, b), which was previously shown to localize to similar interphase

structures¹¹. Furthermore, physical interactions between Blt1–Mid1, Blt1–Cdr2 and Cdr2–Mid1 were detected by co-immunoprecipitation (Fig. 1d). The Cdr2-related kinase Cdr1, which directly phosphorylates and inhibits Wee1 *in vitro*^{12–14}, also co-localized with Blt1 in the nodes (Fig. 1c and Supplementary Fig. 2a). The presence of two components of the mitotic control system, Cdr1 and Cdr2, in nodes during G2 supports the possibility that these structures might function to regulate mitotic entry. To investigate the recruitment of these proteins to the interphase nodes, we monitored green fluorescent protein (GFP)-tagged node proteins in cells deleted for the other components, and found that Cdr2 localized to interphase nodes in *mid1Δ*, *blt1Δ* and *cdr1Δ* cells, but was required for all other proteins to reach the interphase cortical nodes (Fig. 2a and Supplementary Fig. 3a). We determined a hierarchy for recruitment of these proteins to interphase nodes by examining all combinations of GFP-tagged proteins and mutants, including the additional interphase node proteins Klp8 and Gef2 (Supplementary Figs 2, 3 and data not shown). These results indicate that medial cortical nodes are formed by the ordered, Cdr2-dependent assembly of multiple interacting proteins during interphase.

The presence of the mitotic regulators Cdr1 and Cdr2 in the interphase nodes prompted us to examine whether Blt1 also regulates mitotic entry. *blt1Δ* cells had an increased length at division, consistent with a delay in mitotic entry. This phenotype was synthetic with a temperature-sensitive *cdc25^{ts}* mutation (Supplementary Table 1), but not with *cdr1Δ* and *cdr2Δ* mutants, which act through Wee1 (refs 15–17). These genetic data are consistent both with the localization hierarchy described earlier and with a pathway in which Cdr2 functions upstream of Cdr1 and Blt1 to promote mitosis through the negative regulation of Wee1 (Fig. 2b).

To determine whether these nodes might act as sites of Wee1 regulation, we next examined the cellular distribution of Wee1. GFP–Wee1 under the control of the low-level P81nmt1 promoter was found in three cellular structures: the medial cortical nodes, the nucleus and the spindle pole body (Fig. 2c). GFP–Wee1 co-localized with Blt1–mCherry at the medial cortical nodes (Fig. 2d), and with the spindle pole body marker Sid4–mCherry (Supplementary Fig. 4). To confirm this localization pattern for endogenously expressed Wee1, we integrated a triple-GFP tag at the genomic *wee1⁺* locus. Although present at low levels, Wee1–3×GFP was found in the medial cortical nodes and the nucleus (Fig. 2e). Wee1 localization to the medial cortical nodes depended on Cdr2 but not on Cdr1 or Blt1 (Supplementary Fig. 5), reinforcing the importance of Cdr2 in generating these structures. Moreover, a Cdr2(E177A) mutant, which localizes properly but is predicted to abolish Cdr2 kinase activity⁹, disrupted the recruitment of Wee1 to the medial cortex and delayed entry into mitosis (Supplementary Fig. 6), indicating a role for Cdr2 kinase activity in signalling to Wee1. Because Wee1 is present in the same medial cortical nodes as its inhibitory network, we propose that

¹The Rockefeller University, New York, New York 10065, USA. ²Institut Curie, Centre de Recherche and CNRS UMR144, Paris 75248 cedex 05, France.

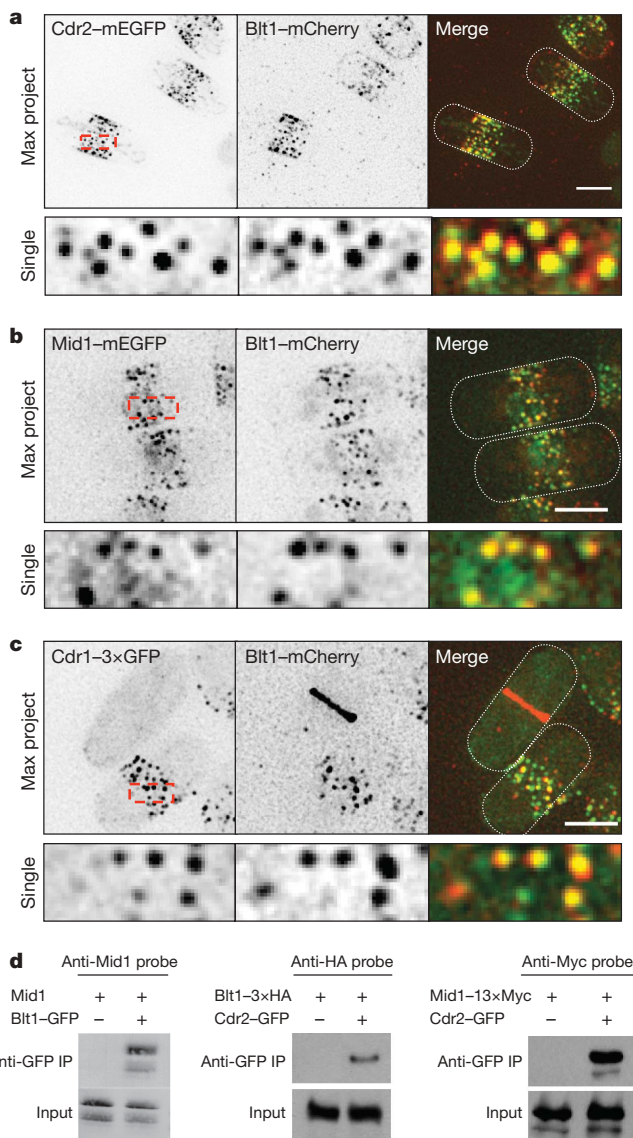


Figure 1 | Interphase medial nodes contain multiple factors that physically interact. **a–c**, Co-localization of Blt1-mCherry and indicated GFP-tagged proteins. Top row images are inverted maximum projections; cells are outlined in white in the merged images. Regions boxed in red are magnified in the bottom rows, which show inverted single focal planes. Max project, maximum projection; mEGFP, monomeric enhanced GFP. Scale bars, 3.5 μ m. **d**, Co-immunoprecipitation (IP) from native cell extracts. Samples were probed with the indicated antibodies.

the interphase nodes control the onset of mitosis through Cdr2-dependent negative regulation of Wee1.

Placement of Cdr2 at the top of the Wee1 regulatory network suggests that mechanisms regulating Cdr2 could act as inputs to the mitotic entry control system. We considered that cell polarity factors positioned at the cell tip might provide spatial cues to limit Cdr2 distribution to the cell middle. To test this possibility, cell polarity factor mutants (*for3* Δ , *mod5* Δ , *orb2*^{ts}, *orb6*^{ts}, *pom1* Δ , *tea1* Δ and *tea4* Δ) were screened for defects in Cdr2 localization. This demonstrated a highly specific role for the protein kinase Pom1, a conserved member of the dual-specificity tyrosine-phosphorylation regulated kinase (DYRK) family of kinases⁵. Cdr2 was restricted to the cell middle in wild-type cells, but localized throughout half of the cell in a *pom1* Δ mutant (Fig. 3a). *pom1* Δ cells grow in a monopolar manner, and the Cdr2 nodes extended from the cell middle out to the non-growing cell end. This strong defect in Cdr2 localization was not

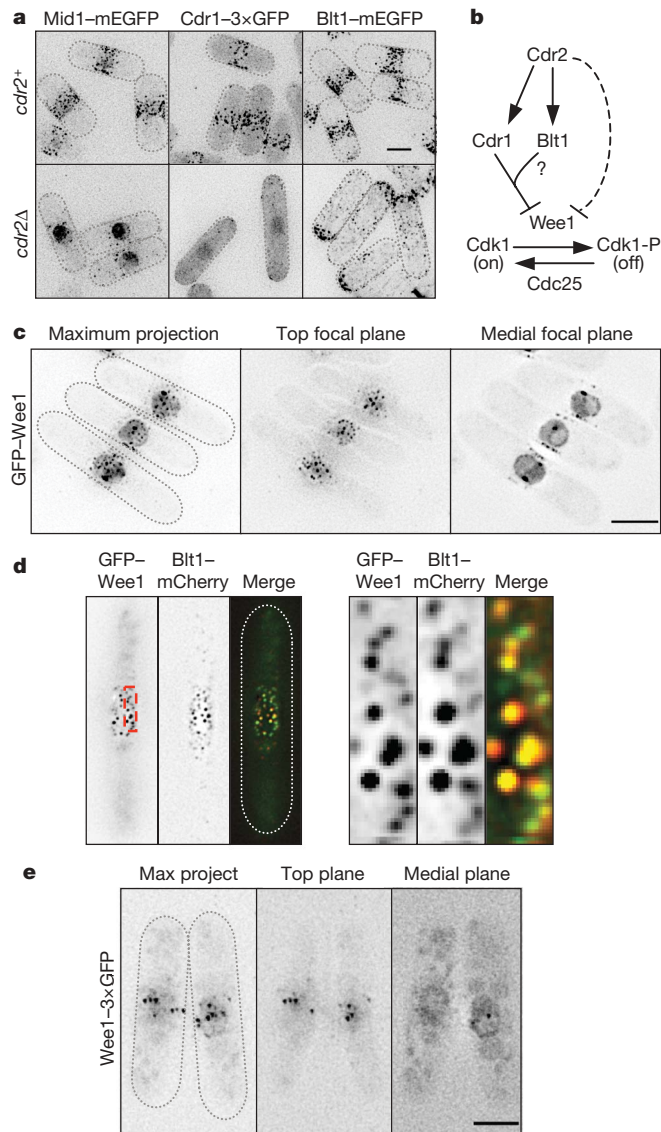


Figure 2 | Interphase nodes are organized by Cdr2 and contain Wee1. **a**, Localization of Mid1-mEGFP, Cdr1-3xGFP and Blt1-mEGFP in wild-type (*cdr2*⁺) and *cdr2* Δ cells. Cells are outlined in black, and images are inverted maximum projections. Scale bar, 3.5 μ m. **b**, Pathway for control of mitotic entry by interphase node proteins on the basis of genetic epistasis and localization analysis. Dashed line indicates potential for direct regulation of Wee1 by Cdr2 (ref 16). Blt1 functions downstream of both Cdr1 and Cdr2, but requires only Cdr2 for proper localization. The molecular function of Blt1 remains unclear, as denoted by the question mark. **c**, Localization of GFP-Wee1 from *S. pombe* strain JM551. Cells are outlined in the left panel. Scale bar, 5 μ m. **d**, Co-localization of GFP-Wee1 and Blt1-mCherry; images are inverted single focal planes. Left panels show a single cell outlined in merged image. The region boxed in red is magnified in the right panels. **e**, Localization of Wee1-3xGFP expressed from the endogenous promoter and chromosomal locus. Scale bar, 3.5 μ m.

observed in other monopolar or cell polarity mutants (Supplementary Fig. 7), although a more minor defect was found in *tea1* Δ and *tea4* Δ mutants, which fail to localize Pom1 to cell ends^{5,18}. Interphase node components downstream of Cdr2 also localized to the cell end in *pom1* Δ mutants (Supplementary Fig. 8; previously shown for Mid1 (refs 19, 20)).

These results indicate that Pom1 provides inhibitory signals that confine Cdr2 to the cell middle. Pom1-dependent signals control Cdr2 phosphorylation, as shown by the retardation of Cdr2-haemagglutinin (HA) migration in protein gels owing to phosphorylation after overexpression of the Pom1 protein kinase (Fig. 3b). Consistently,

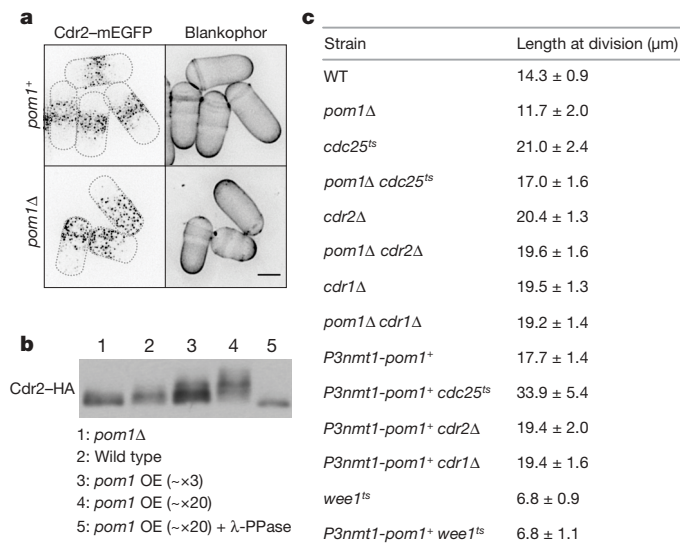


Figure 3 | Pom1 regulates Cdr2 and mitotic entry. **a**, Localization of Cdr2-mEGFP in wild-type (*pom1*⁺) and *pom1Δ* cells that are also stained with Blankophor, a cell wall dye that brightly stains growing cell ends. Images are inverted maximum projections. Scale bar, 3.5 μm. **b**, Migration of Cdr2-3×HA from wild-type, *pom1Δ* and *pom1*-overexpressing (OE) cells. Whole-cell extracts were separated by SDS-PAGE, transferred to nitrocellulose, and probed with an anti-HA antibody. Levels of *pom1*⁺ overexpression were quantified by western blot (Supplementary Fig. 11a, b). λ-PPase, λ-phosphatase. **c**, Length of dividing, septated cells of the indicated genotypes (mean ± s.d.; *n* > 50 for each value). *wee1*^{ts} strains were shifted to 36.5 °C for 6 h before measurements; all other strains were measured at 25 °C. *P3nmt1-pom1*⁺ cells were grown in the presence of thiamine, which increases Pom1 levels approximately threefold compared to endogenous levels (Supplementary Fig. 11a).

Cdr2-HA migrated faster after the deletion of *pom1*⁺ (Fig. 3b and Supplementary Fig. 9). These results indicate that Pom1 may act directly on Cdr2 in cells, although an indirect effect through other protein kinases remains possible. To investigate potential functional roles for this regulation, we examined the length of *pom1Δ* cells at division. These divided at a smaller size than wild-type cells, indicating premature entry into mitosis (Fig. 3c). *pom1Δ* suppressed the elongated cell-length phenotype of a *cdc25ts* mutant, but had little effect when combined with *cdr2Δ* and *cdr1Δ*. Conversely, increased *pom1*⁺ expression generated a delay in mitotic entry that was synthetic with *cdc25ts* but not with *cdr2Δ*, *cdr1Δ* or *wee1ts*. This mitotic delay correlated with increased phosphorylation of Cdk1 on Tyr15 (Supplementary Fig. 10), the target of Wee1 kinase². These data indicate that Pom1 functions in a dose-dependent manner to delay entry into mitosis through negative regulation of the Cdr2-Cdr1-Wee1 pathway. The *pom1Δ* size phenotype is not as severe as *wee1*⁺ deletion, indicating that further Wee1 regulatory mechanisms are likely to be operating. We conclude that Pom1 is a potential functional link between polarized cell growth and mitotic entry by regulating these two processes.

Pom1 forms a polar gradient that peaks at cell ends¹⁹. Using quantitative microscopy on cells expressing Pom1-GFP, we examined the changes in Pom1 distribution as cells elongate and progress through G2 (Fig. 4a). Fluorescence-intensity measurements revealed the presence of Pom1 in the middle of small cells. As cell length increased, Pom1 levels in the cell middle showed a progressive decrease (Fig. 4b); this size-dependent decrease was confirmed in a larger number of cells (Supplementary Fig. 12). Thus, Cdr2 localized at the cell middle encounters progressively lower levels of inhibitory Pom1 as the cell elongates and progresses through G2. Given these results, we propose a model for how the relationship between Cdr2 medial nodes and a Pom1 inhibitory gradient might relay information about cell size to

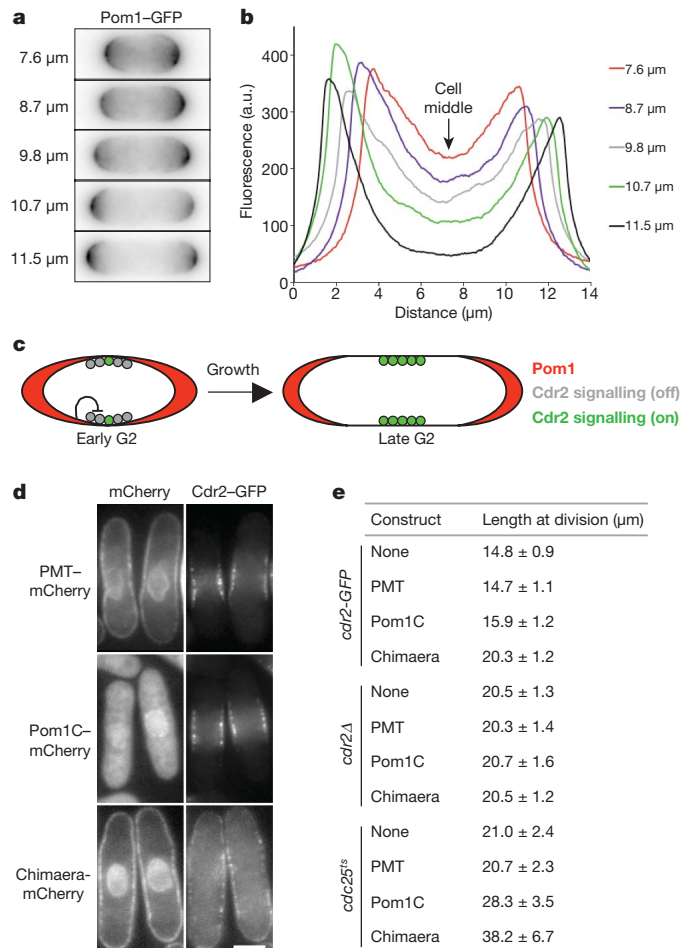


Figure 4 | Pom1 levels at the cell centre control Cdr2-dependent entry into mitosis. **a**, Localization of Pom1-GFP in cells of indicated size. Images are inverted sum projections. **b**, Fluorescence intensity quantifications for cells from **a**. a.u., arbitrary units. **c**, Model for regulation of mitotic entry at Cdr2 nodes by Pom1 gradient. **d**, Left: localization of PMT, Pom1C and chimaera (PMT-Pom1C fusion). Constructs are tagged with mCherry. Right: localization of Cdr2-GFP in cells expressing the indicated constructs. Images are single focal planes, and the same cells are shown for mCherry and Cdr2-GFP images. Scale bar, 3.5 μm. **e**, Length of dividing, septated cells expressing the indicated constructs in *cdr2-GFP*, *cdr2Δ* or *cdc25-22* background, as indicated (mean ± s.d.; *n* > 50 for each value).

regulate mitotic entry (Fig. 4c). In the model, small cells in early G2 contain sufficient levels of Pom1 at the cell middle to negatively regulate Cdr2 function and prevent entry into mitosis. After polarized growth during G2, Pom1 levels at the cell centre decrease, allowing Cdr2 nodes to promote mitosis by inhibiting Wee1. In support of cell size-dependent regulation by Pom1, Cdr2-HA from small cells (*wee1ts*) migrates slower in protein gels than Cdr2-HA from large cells (*cdc25ts*) (Supplementary Fig. 9c).

This model predicts that cell cycle regulation at medial nodes responds to changes in Pom1 localization. To test this prediction, we first examined *tea1Δ* cells, in which Pom1 is not concentrated at cell ends but does retain full kinase activity²¹. This altered distribution within the cell should allow Pom1 more access to Cdr2 at the cell middle. Consistent with the model, *tea1Δ* cells showed a delay in mitotic entry, and this phenotype was additive with the *cdc25ts* mutation but not with the *cdr2Δ* or *pom1Δ* mutants (Supplementary Table 2). As a more direct test, we ectopically localized Pom1 kinase activity throughout the cortex by fusing the Pom1 carboxy-terminal kinase domain (Pom1C) to a plasma membrane targeting domain (PMT). This chimaera was efficiently targeted throughout the cortex and disrupted Cdr2 localization in the cell middle (Fig. 4d).

Expression of the chimaera generated a delay in mitotic entry that was equivalent to the deletion of *cdt2*⁺, whereas the chimaera did not affect *cdt2Δ* cells (Fig. 4e). Furthermore, expression of the chimaera in *cdc25*^{ts} cells led to a strong mitotic delay. We conclude that the levels of Pom1 at the cell middle regulate entry into mitosis through the negative regulation of the Cdr2 pathway.

By exploiting the simple geometry and genetic tractability of fission yeast, we have revealed a regulatory system coordinating cell size and division by means of a polar gradient that inhibits the function of cortical structures in the cell middle. A central component of this mechanism is the conserved kinase Pom1, which acts as a molecular link between cell growth and mitotic entry through a Cdr2-Cdr1-Wee1-Cdk1 pathway. We propose that the Pom1 polar gradient provides information about cell size and geometry for the Cdk1 mitotic regulatory system. An important next step will be to define the biochemical nature of interactions within this signalling network, which will lead to the generation of quantitative models for cell cycle regulation at interphase nodes. Such models have already provided insight into the mechanism of cytokinetic ring assembly by post-mitotic cortical nodes²².

The ability of diffusible morphogen gradients to coordinate size and geometry has been much studied in multicellular developmental systems. However, recent work has demonstrated that gradients also operate within cells to control a growing number of processes, including GTPase signalling²³, mitotic spindle assembly²⁴, and the mitotic entry system that we have described here. Such intracellular gradients operate on a smaller spatial scale, but are able to generate information about cell size and shape²³. Thus, the use of gradients to generate spatial information represents a widely conserved mechanism operating at a range of scales in biology.

METHODS SUMMARY

Standard *S. pombe* media and methods were used²⁵. Strains are listed in Supplementary Table 3. Gene deletion and tagging were performed by PCR and homologous recombination²⁶. Cells for length measurements were stained with Blankophor (MP Biochemicals), imaged in Metamorph (MDS Analytical Technologies) using an epifluorescence microscope (Axioplan 2, Carl Zeiss, Inc.) equipped with a CoolSNAP HQ camera (Roper Scientific), and analysed in ImageJ (National Institutes of Health). More than 50 septated cells were measured for each strain.

For protein localization, cells were imaged on agar pads or in liquid media under a coverslip with 0.4 or 0.5 μm step size using a DeltaVision Spectris (Applied Precision) comprised of an Olympus IX71 wide-field inverted fluorescence microscope, an Olympus UPlanSapo ×100, numerical aperture 1.4, oil immersion objective, and a Photometrics CoolSNAP HQ camera (Roper Scientific). Images were captured and processed by iterative constrained deconvolution using SoftWoRx (Applied Precision), and analysed in ImageJ. For quantitative Pom1 localization, Pom1-GFP cells (JM414) were imaged in 14 Z-sections of 0.4 μm each, and a sum projection of each single cell was compressed into a one-dimensional line along the long axis of the cell using ImageJ. The fluorescence intensity profile of this line was plotted, and the background fluorescence from cells not expressing GFP was subtracted. The minimum fluorescence value in the centre of each cell was identified from these background-subtracted profiles to generate the mean values in Supplementary Fig. 12.

For co-immunoprecipitation, cells were washed in IP buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.2% NP-40, 1 mM PMSF and Complete protease inhibitor (Roche)) and lysed by bead-beating. Extracts were cleared by centrifugation at 16,000g and incubated with Dynabeads coated with anti-GFP²⁷. Precipitates were washed in IP buffer, separated by SDS-PAGE, and analysed by western blotting. Whole-cell lysates were prepared as described²⁸.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 31 January; accepted 21 April 2009.
Published online 27 May 2009.

1. Mitchison, J. M. Growth during the cell cycle. *Int. Rev. Cytol.* **226**, 165–258 (2003).
2. Jorgensen, P. & Tyers, M. How cells coordinate growth and division. *Curr. Biol.* **14**, R1014–R1027 (2004).

3. Rupes, I. Checking cell size in yeast. *Trends Genet.* **18**, 479–485 (2002).
4. Dolznig, H., Grebien, F., Sauer, T., Beug, H. & Müllner, E. W. Evidence for a size-sensing mechanism in animal cells. *Nature Cell Biol.* **6**, 899–905 (2004).
5. Bähler, J. & Pringle, J. R. Pom1p, a fission yeast protein kinase that provides positional information for both polarized growth and cytokinesis. *Genes Dev.* **12**, 1356–1370 (1998).
6. Nurse, P., Thuriaux, P. & Nasmyth, K. Genetic control of the cell division cycle in the fission yeast *Schizosaccharomyces pombe*. *Mol. Gen. Genet.* **146**, 167–178 (1976).
7. Nurse, P. Genetic control of cell size at cell division in yeast. *Nature* **256**, 547–551 (1975).
8. Ubersax, J. A. et al. Targets of the cyclin-dependent kinase Cdk1. *Nature* **425**, 859–864 (2003).
9. Morrell, J. L., Nichols, C. B. & Gould, K. L. The GIN4 family kinase, Cdr2p, acts independently of septins in fission yeast. *J. Cell Sci.* **117**, 5293–5302 (2004).
10. Wu, J. Q., Kuhn, J. R., Kovar, D. R. & Pollard, T. D. Spatial and temporal pathway for assembly and constriction of the contractile ring in fission yeast cytokinesis. *Dev. Cell* **5**, 723–734 (2003).
11. Paoletti, A. & Chang, F. Analysis of mid1p, a protein required for placement of the cell division site, reveals a link between the nucleus and the cell surface in fission yeast. *Mol. Biol. Cell* **11**, 2757–2773 (2000).
12. Wu, L. & Russell, P. Nim1 kinase promotes mitosis by inactivating Wee1 tyrosine kinase. *Nature* **363**, 738–741 (1993).
13. Parker, L. L., Walter, S. A., Young, P. G. & Piwnicka-Worms, H. Phosphorylation and inactivation of the mitotic inhibitor Wee1 by the *nim1/cdr1* kinase. *Nature* **363**, 736–738 (1993).
14. Coleman, T. R., Tang, Z. & Dunphy, W. G. Negative regulation of the wee1 protein kinase by direct action of the *nim1/cdr1* mitotic inducer. *Cell* **72**, 919–929 (1993).
15. Breeding, C. S. et al. The *cdt2*⁺ gene encodes a regulator of G2/M progression and cytokinesis in *Schizosaccharomyces pombe*. *Mol. Biol. Cell* **9**, 3399–3415 (1998).
16. Kanoh, J. & Russell, P. The protein kinase Cdr2, related to Nim1/Cdr1 mitotic inducer, regulates the onset of mitosis in fission yeast. *Mol. Biol. Cell* **9**, 3321–3334 (1998).
17. Russell, P. & Nurse, P. The mitotic inducer *nim1*⁺ functions in a regulatory network of protein kinase homologs controlling the initiation of mitosis. *Cell* **49**, 569–576 (1987).
18. Martin, S. G., McDonald, W. H., Yates, J. R. & Chang, F. Tea4p links microtubule plus ends with the formin for3p in the establishment of cell polarity. *Dev. Cell* **8**, 479–491 (2005).
19. Padte, N. N., Martin, S. G., Howard, M. & Chang, F. The cell-end factor pom1p inhibits mid1p in specification of the cell division plane in fission yeast. *Curr. Biol.* **16**, 2480–2487 (2006).
20. Celton-Morizur, S., Racine, V., Sibarita, J. B. & Paoletti, A. Pom1 kinase links division plane position to cell polarity by regulating Mid1p cortical distribution. *J. Cell Sci.* **119**, 4710–4718 (2006).
21. Bähler, J. & Nurse, P. Fission yeast Pom1p kinase activity is cell cycle regulated and essential for cellular symmetry during growth and division. *EMBO J.* **20**, 1064–1073 (2001).
22. Vavylonis, D., Wu, J. Q., Hao, S., O'Shaughnessy, B. & Pollard, T. D. Assembly mechanism of the contractile ring for cytokinesis by fission yeast. *Science* **319**, 97–100 (2008).
23. Meyers, J., Craig, J. & Odde, D. J. Potential for control of signaling pathways via cell size and shape. *Curr. Biol.* **16**, 1685–1693 (2006).
24. Bastiaens, P., Caudron, M., Niethammer, P. & Karsenti, E. Gradients in the self-organization of the mitotic spindle. *Trends Cell Biol.* **16**, 125–134 (2006).
25. Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).
26. Bähler, J. et al. Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14**, 943–951 (1998).
27. Cristea, I. M., Williams, R., Chait, B. T. & Rout, M. P. Fluorescent proteins as proteomic probes. *Mol. Cell. Proteomics* **4**, 1933–1941 (2005).
28. Sreenivasan, A. & Kellogg, D. The Elm1 kinase functions in a mitotic signaling network in budding yeast. *Mol. Cell. Biol.* **19**, 7983–7994 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Nurse laboratory for discussions and critical reading of the manuscript, A. Puszyńska for technical assistance, The Rockefeller University Bio-Imaging Resource Center for assistance with microscopy, The Rockefeller University Proteomics Resource Center for mass spectrometry, K. Gould, J.-Q. Wu, T. Pollard, J. Bähler, F. Chang and K. Sawin for yeast strains, plasmids and antibodies, and S. Martin for sharing unpublished data and for discussions. J.B.M. was supported by a postdoctoral fellowship from the American Cancer Society (PF-07-129-01-MBC); A.P. by funding from ANR (BLAN06-3_135468), ARC (no. 4934) and FRM (INE2007110973); and P.N. by the Breast Cancer Research Foundation and The Rockefeller University.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.B.M. (jmoseley@rockefeller.edu).

METHODS

Yeast methods. Yeast strains were generated by tetrad dissection, when applicable. For cell-length measurements, cells were grown at 25 °C in EMM media lacking supplements. Cells measured for Supplementary Fig. 6d were grown in EMM+4S media. Strains measured in Fig. 3c were: JM366, JM185, JM14, JM618, JM554, JM616, JM555, JM492, JM717, JM787, JM805, JM887, JM228 and JM786. Strains measured in Fig. 4e were: AP2458, AP2542, AP2541, AP2540, JM554, AP2545, AP2544, AP2543, JM14, AP2587, AP2468 and AP2588. Strains measured in Supplementary Fig. 6d were: JM823, JM822 and JM826. Strains measured in Supplementary Table 1 were: JM366, JM206, JM14, JM357, JM554, JM888, JM555, JM889 and JM625. Strains measured in Supplementary Table 2 were: JM366, JM885, JM14, JM619, JM554, JM614 and JM886. Statistical analyses of genetic interactions are listed in Supplementary Table 4 ($\alpha = 0.05$). The only GFP-tagged strains for which we noticed an altered cell length at division were *cdr1-3GFP* ($11.4 \pm 1.3 \mu\text{m}$) and *wee1-3GFP* ($16.5 \pm 1.6 \mu\text{m}$); these phenotypes suggest increased activity of Cdr1 and Wee1, respectively, in these strains. *P81nmt1-GFP-wee1⁺* strains were grown in the presence of thiamine, and expression was induced by removal of thiamine for 20 h before imaging. PMT encodes amino acids 500–920 of Mid1 and localizes to the cortex independently of Cdr2 (Supplementary Fig. 13c). Pom1C encodes amino acids 591–1087 of Pom1. PMT, Pom1C and the PMT–Pom1C chimaera were subcloned into the previously described pAP140 (ref. 29), modified to contain the mCherry coding sequence (from pKS390, a gift from K. Sawin). To generate pJM353, the *cdr2⁺* promoter, *cdr2⁺* coding sequence, GFP coding sequence and *cdr2⁺* terminator were PCR-amplified and subcloned into the integration plasmid pJK148. The mutation E177A was introduced by site-directed mutagenesis of pJM353 using Quikchange IIXL kit (Stratagene) to generate pJM385. For pJM394, the *cdr2⁺* locus including promoter (–800) and terminator (+1,000) was PCR-amplified and subcloned into integration plasmid pJM391 (gift from D. Coudreuse). The

mutation E177A was introduced by site-directed mutagenesis of pJM394 using Quikchange IIXL kit (Stratagene) to generate pJM397.

Biochemistry. For mass spectrometry, protein complexes were isolated from strains JM10 and JM151 using described methods²⁷. Polypeptides were separated by SDS–PAGE followed by Coomassie staining. The gel bands were excised, diced and treated with 10 mM dithiothreitol in 0.1 M ammonium bicarbonate for protein reduction. Free cysteine residues were alkylated with freshly made 55 mM iodoacetamide in 0.1 M ammonium bicarbonate. Tryptic digestion was started with the addition of $25 \text{ ng } \mu\text{l}^{-1}$ Sequence Grade Modified Trypsin (Promega) in ammonium bicarbonate buffer. The protein was digested for at least 16 h at 37 °C. The resultant tryptic peptide mixture was separated by gradient elution with the Dionex nano-HPLC system that was interfaced with a Thermo LTQ mass spectrometer (MS). Automated data-dependent acquisition was used to acquire MS/MS data. The acquired MS/MS spectra were converted to a MASCOT acceptable format and searched by the Mascot search engine to identify proteins from primary sequence databases.

Western blots were probed with antibodies anti-HA (F-7, Santa Cruz Biotechnology, Inc.), anti-Myc (9E10, Santa Cruz Biotechnology, Inc.), anti-GFP (generated as described²⁷), anti-Mid1 (ref. 29), anti-Cdc2 (phospho-Tyr15) (9111, Cell Signaling Technology), anti-PSTAIR (sc-53, Santa Cruz Biotechnology, Inc.), and anti-Pom1 (gift from J. Bähler). Quantification of Pom1 levels is shown in Supplementary Fig. 11. Band intensities of scanned blots were quantified using ImageJ. The integrated intensity of a fixed area was measured, and background levels were subtracted.

29. Celton-Morizur, S., Bordes, N., Fraissier, V., Tran, P. T. & Paoletti, A. C-terminal anchoring of mid1p to membranes stabilizes cytokinetic ring position in early mitosis in fission yeast. *Mol. Cell. Biol.* **24**, 10621–10635 (2004).

Chaperone-mediated pathway of proteasome regulatory particle assembly

Jeroen Roelofs¹, Soyeon Park¹, Wilhelm Haas¹, Geng Tian¹, Fiona E. McAllister¹, Ying Huo¹, Byung-Hoon Lee¹, Fan Zhang², Yigong Shi², Steven P. Gygi¹ & Daniel Finley¹

The proteasome is a protease that controls diverse processes in eukaryotic cells. Its regulatory particle (RP) initiates the degradation of ubiquitin–protein conjugates by unfolding the substrate and translocating it into the proteasome core particle (CP) to be degraded¹. The RP has 19 subunits, and their pathway of assembly is not understood. Here we show that in the yeast *Saccharomyces cerevisiae* three proteins are found associated with RP but not with the RP–CP holoenzyme: Nas6, Rpn14 and Hsm3. Mutations in the corresponding genes confer proteasome loss-of-function phenotypes, despite their virtual absence from the holoenzyme. These effects result from deficient RP assembly. Thus, Nas6, Rpn14 and Hsm3 are RP chaperones. The RP contains six ATPases—the Rpt proteins—and each RP chaperone binds to the carboxy-terminal domain of a specific Rpt. We show in an accompanying study² that RP assembly is templated through the Rpt C termini, apparently by their insertion into binding pockets in the CP. Thus, RP chaperones may regulate proteasome assembly by directly restricting the accessibility of Rpt C termini to the CP. In addition, competition between the RP chaperones and the CP for Rpt engagement may explain the release of RP chaperones as proteasomes mature.

Affinity purification has allowed identification of proteins that associate with proteasomes but are not true subunits^{1,3–5}. Nas6, one such protein⁴, is the apparent homologue of gankyrin (PSMD10, also known as p28), a liver oncoprotein. Gankyrin is thought to interact with retinoblastoma protein (RB), MDM2, CDK4 and the proteasome⁶. Using recombinant glutathione S-transferase (GST)–Nas6 we purified all 19 RP subunits from yeast lysates (Supplementary Table 1). However, to our surprise, no CP subunit was recovered. Similarly, affinity purified samples from RP-tagged strains contained Nas6 whereas those from CP-tagged strains did not (Fig. 1a). Thus, Nas6 may bind a subpopulation of RP that is not associated with CP (hereafter ‘free RP’). The specific association of Nas6 with the RP was validated through additional experiments (Supplementary Fig. 1).

Mass spectrometric analysis of immunoprecipitates of haemagglutinin (HA)-tagged Nas6 revealed two other proteins that bind specifically to free RP, Hsm3 and Rpn14. Hsm3 was recently identified as a proteasome-interacting protein⁷. Hsm3 co-purified with the RP but not CP (Fig. 1b), confirming that it is a second RP-specific component. Although we cannot exclude the possibility that a fraction of these proteins shows bona fide proteasome holoenzyme association, their previous assignment as proteasome subunits may reflect that proteasome subcomplexes are often found in holoenzyme preparations. The mammalian homologue of Rpn14, PAAF1, was recently suggested to bind free RP⁸. Likewise, Rpn14 associated specifically with free RP (Fig. 1c). Hsm3 differed from Rpn14 and Nas6 in associating most strongly with only a subset of RP subunits, potentially representing an RP assembly intermediate (Fig. 1d).

The identification of multiple free RP-binding proteins indicates a common function distinct from those of known proteasome-associated proteins. Free RP might be actively sequestered from CP by an inhibitor of the proteasome, a function proposed for PAAF1 (refs 8 and 9). Alternatively, free RP might provide a non-proteolytic pathway of RP function, suggested for transcriptional control and for PAAF1 in particular^{9,10}. Third, free RP-binding proteins might act as RP assembly factors. The mechanisms involved in RP assembly^{11–13} are largely unknown. In contrast, CP assembly has been studied intensively and involves specific chaperones¹⁴.

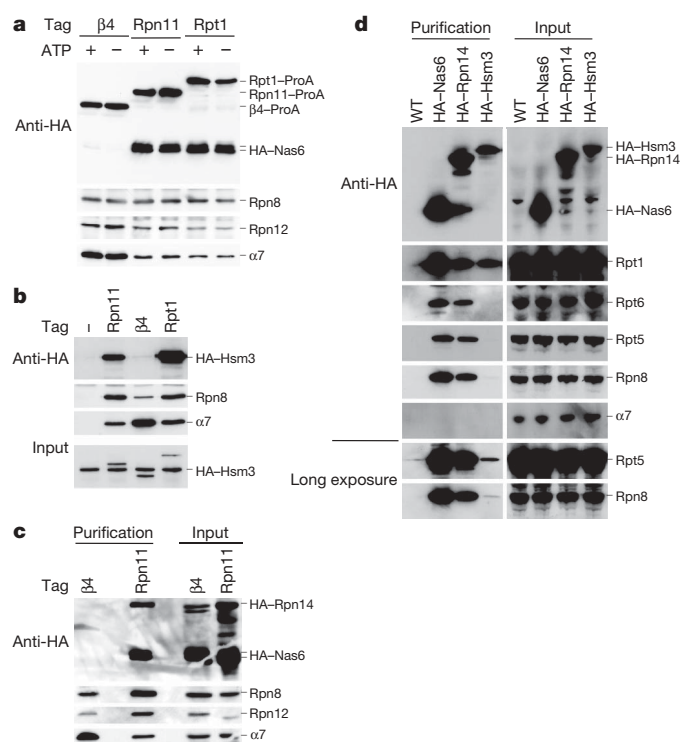


Figure 1 | Nas6, Hsm3 and Rpn14 bind to free RP. **a**, Proteasomes were affinity-purified from strains expressing HA-tagged Nas6. Samples were resolved by SDS–PAGE and immunoblotted for Nas6 (anti-HA), RP (anti-Rpn8 and anti-Rpn12) and CP (anti- α 7). **b**, Proteasomes were affinity-purified from strains expressing HA-tagged Hsm3, eluted and analysed for the presence of Hsm3, RP and CP. **c**, Proteasomes were affinity-purified from strains expressing HA-tagged Nas6 and Rpn14. After elution, samples were analysed as in **a**. **d**, HA-specific immunoprecipitations were tested for RP and CP by immunoblotting. ProA, protein A; WT, wild type. α 7 and β 4 are CP subunits.

¹Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA. ²Department of Molecular Biology, Lewis Thomas Laboratory, Princeton University, Princeton, New Jersey 08544, USA.

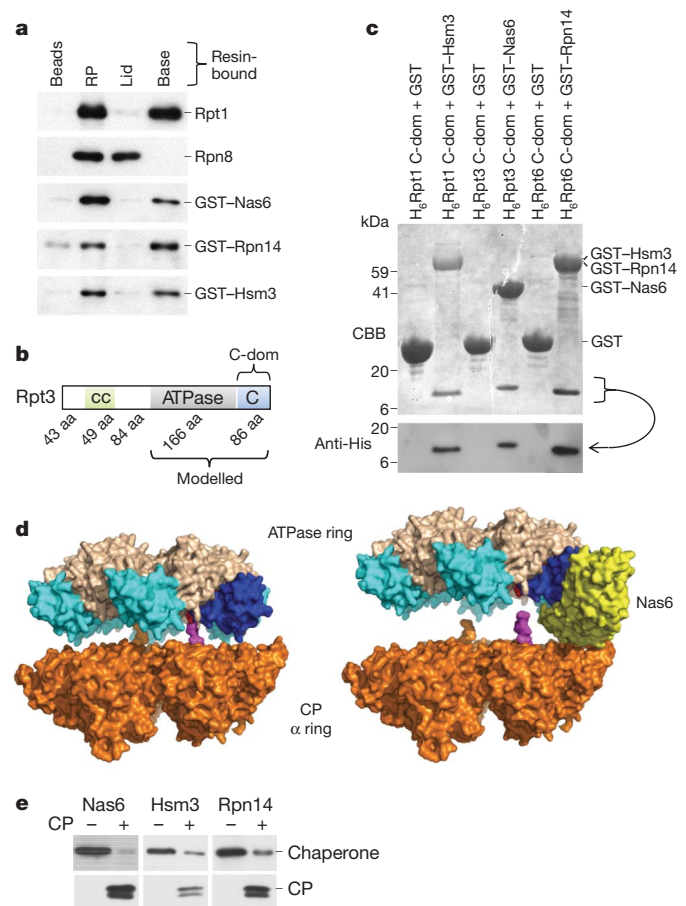


Figure 3 | RP chaperones bind to the C-domains of Rpt proteins in proximity to the CP. **a**, RP, lid and base bound to IgG beads or untreated IgG beads were incubated with GST-Nas6, GST-Rpn14 or GST-Hsm3 purified from *E. coli*. Beads were analysed for bound proteins. **b**, Domain composition of Rpt3; other Rpt proteins have a similar architecture. aa, amino acids; C, C-domain²⁰; cc, predicted coiled-coil region. **c**, His-tagged C-domains were co-expressed with GST or GST-tagged Nas6, Rpn14 or Hsm3 in *E. coli*. Glutathione-Sepharose-purified samples were immunoblotted using His-tag antibody followed by Coomassie blue staining (CBB). **d**, Modelling of the interaction of the Rpt ring with CP in the presence (right) and absence (left) of Nas6: Nas6 binding to the ATPase ring appears to block the C-terminal tail of Rpt3 from docking into the CP. The CP α ring (orange)²⁴ was combined with the PAN hexameric ATPase²³ (ATPase domain, grey; C-domains, blue). The Nas6-Rpt3 structure²¹ was mapped onto the ATPase ring by aligning the Rpt3 structure with the PAN C-domain (dark blue; C terminus, red), with root mean squared deviation of 1.1 Å over 283 atoms. The last 12 residues of the proteasome activator PA26 C terminus¹⁷ (magenta) are docked into CP^{17–19}. The length corresponds to that of Rpt3 C-terminal tail that is absent from the Nas6-Rpt3 structure. **e**, Affinity-purified, resin-bound base was incubated with indicated RP chaperone, washed and then loaded with purified CP. Resin-bound proteins were eluted and analysed by immunoblotting.

The direct interaction of all three RP chaperones with distinct Rpt proteins is consistent with their ability to function in each other's absence (Fig. 2a). These results also explain why Hsm3 but not Nas6 or Rpn14 associates with BP1; their binding partners are absent from BP1 (ref. 2). In support of the functional significance of the Hsm3-Rpt1 interaction and of BP1, we found that overexpressed Rpt1 rescued phenotypes of *rpn4Δ hsm3Δ*, *nas6Δ hsm3Δ* and *rpn14Δ hsm3Δ* mutants but not a *nas6Δ rpn14Δ* mutant (Supplementary Fig. 8). Thus, Rpt1 overexpression relieves the deleterious effects of the absence of Hsm3. Taken together, the data indicate that Hsm3 chaperones Rpt1 or an Rpt1-containing complex.

The Rpt C termini mediate the interaction of RP with the CP^{17–19}, and are moreover critical for base assembly², indicating that assembly

is templated on the CP. In contrast to the RP-CP interaction, binding of RP chaperones to C-domains of their cognate ATPases does not involve the extreme C-terminal residues (Supplementary Fig. 7). We propose that, despite non-overlapping binding sites, RP chaperones compete with CP for binding to Rpt proteins, thus controlling the exposure of Rpt C termini to the CP.

To study the effect of chaperone binding to the Rpt on RP-CP interactions, we modelled the Rpt3-Nas6 structure^{21,22} into an ATPase ring based on recent structural studies of the proteasome-activating nucleotidase (PAN), a closely related ATPase from the archaeon *Methanocaldococcus jannaschii*²³. The ATPase ring was lowered onto the known structure²⁴ of the CP α ring. The modelling predicts that Nas6 physically occludes the formation of proper CP-RP contacts (Fig. 3d), because much of Nas6 projects from the RP towards the CP, to prevent close RP-CP apposition.

A prediction of the physical occlusion model is that the CP should be able to expel RP chaperones from the RP. To test this, a base sample loaded with chaperones was immobilized on resin. Application of purified CP to the resin resulted in efficient elution of the RP chaperones (Fig. 3e). This experiment may mimic key steps in proteasome assembly. Moreover, the direct competition between the RP chaperones and the CP shown here appears to explain the absence of RP chaperones from mature proteasomes. Additional tests of the model are described in the accompanying study².

The yeast protein Nas6 and the human oncoprotein gankyrin are closely related by structure, and bind orthologous Rpt proteins^{21,22}. In contrast to Nas6, however, gankyrin is thought to be a major proteasome component⁶. To test for functional conservation between Nas6 and gankyrin, we expressed gankyrin in a *nas6Δ rpn14Δ* strain. Gankyrin rescued this strain's temperature sensitivity, showing strong evolutionary conservation (Fig. 4a). Gankyrin expressed in yeast associated specifically with free RP (Fig. 4b). Binding of gankyrin to RP was reduced when Nas6 was overexpressed, indicating that gankyrin and Nas6 bind the same RP surface in yeast (Fig. 4b). Consistently, gankyrin bound the C-domain of human RPT3 (PSMC4) but not those of other Rpt proteins²² (Fig. 4c). Moreover, preparations of human proteasomes also showed specific association of endogenous gankyrin with free RP and not holoenzyme (Fig. 4d). Considering the low levels of free RP in the preparation, the specificity of gankyrin for free RP is appreciable. Taken together, these data indicate that the functions of gankyrin and Nas6 have indeed been conserved in evolution.

Although BLAST searches did not reveal a human orthologue of Hsm3, the only remaining mammalian proteasome subunit not found in *S. cerevisiae*, S5b (PSMD5), contains ARM repeats, like Hsm3. S5b can apparently bind to a human RPN1-RPT1-RPT2 complex²⁵, further suggesting that these two proteins are orthologous. When tested against a panel of human Rpt proteins, S5b bound specifically to human RPT1 (PSMC2), as predicted from the specificity of Hsm3 for yeast Rpt1 (Fig. 4e). To test for BP1 in human cells, we analysed HeLa cells lysates on native gels, and subjected gel slices to mass spectrometry. The results showed that S5b co-migrated with human RPN1, RPT1 and RPT2 (PSMD2, PSMC2 and PSMC1, respectively) in a complex that was clearly resolved from proteasomes and RP (Fig. 4f). The existence of BP1 in human cells was confirmed through immunoblotting (Fig. 4g).

Here we identify three factors that assist in assembly of the proteasome base through a common mechanism of action: they associate with C-domains of Rpt proteins, thus negatively regulating interaction of the C-domain with specific pockets within the CP α ring. Modelling indicates that RP chaperones prevent CP interaction by steric occlusion, and as a result compete with the CP for occupancy of C-domains. This model was tested in three ways: free CP releases RP chaperones from the base complex (Fig. 3e); single amino acid deletions of Rpt C termini fail to expel their cognate RP chaperones², indicating that proper docking of the C terminus is needed to expel the chaperone; and insertions of as little as one residue proximal to

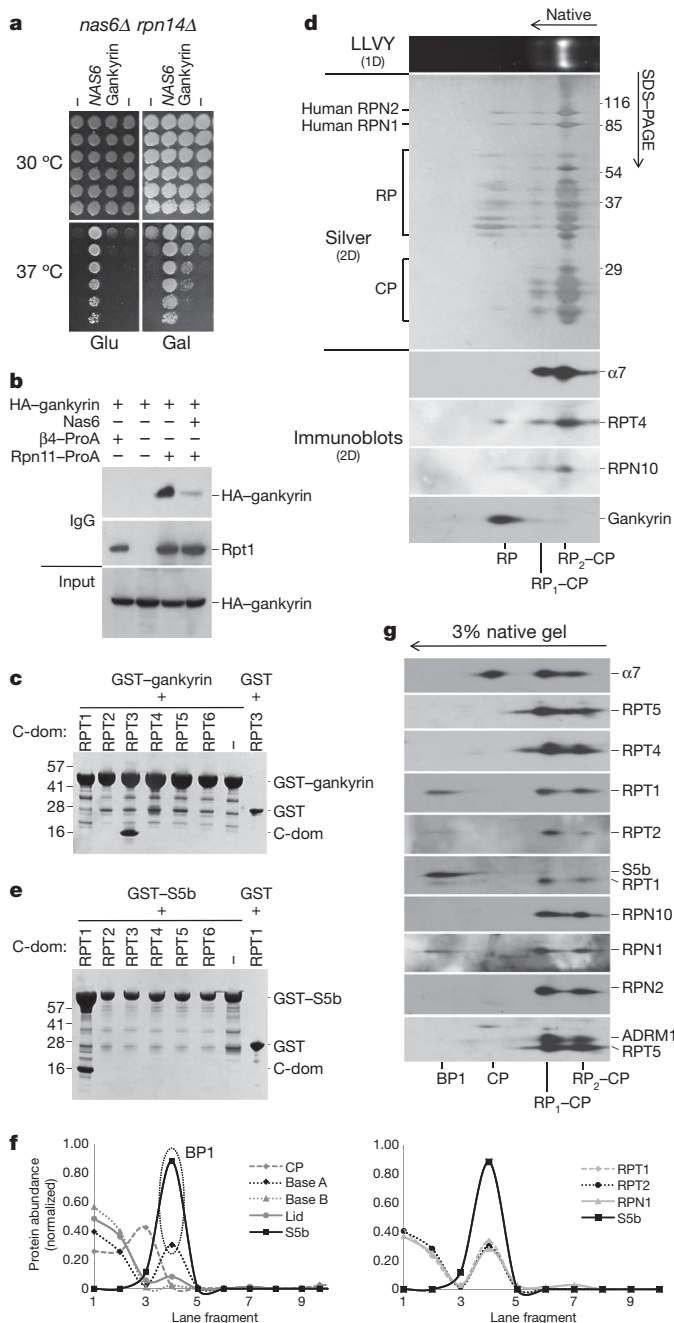


Figure 4 | Evolutionary conservation of the proteasome assembly pathway. **a**, *nas6Δ rpn14Δ* strains expressing *NAS6* or gankyrin from a *GAL* promoter were spotted onto plates containing rich medium with glucose or galactose in fourfold dilutions, and incubated at the indicated temperatures. **b**, Proteasomes were purified from yeast strains expressing HA-tagged gankyrin and analysed by immunoblotting. IgG, immunoglobulin G. **c**, His-tagged C-domains of the indicated human Rpt proteins were co-expressed in *E. coli* with GST-tagged gankyrin. Glutathione–Sephacel-purified samples were separated by SDS–PAGE, and stained with Coomassie blue. Band assignments were confirmed by immunoblotting (data not shown). **d**, Human proteasomes purified via HTBH-tagged Rpn11 (ref. 5) were resolved by two-dimensional (2D) native SDS–PAGE gel electrophoresis. Proteins were silver-stained or analysed by immunoblotting. **e**, As in **c**, only GST–gankyrin was replaced by GST–S5b. **f**, HeLa cell lysate was resolved on Blue native gels. Fractions were analysed by mass spectrometry, and protein abundance was estimated using spectral counting. On the basis of their abundance profiles, most proteasome subunits can be divided into four groups (left): CP (human (h)α1–7 and hβ1–7), lid (hRPN3, 5, 6, 7, 8, 9, 10, 11), base B (hRPT3, 4, 5, 6 and hRPN2) and base A (hRPT1, 2 and hRPN1). Interestingly, S5b is absent from proteasomes (fraction 1 and 2), but abundant in fraction 4. Right panel, S5b and individual profiles of Base A members. **g**, HeLa cell lysates from **e** were separated by two-dimensional native-SDS–PAGE and analysed by immunoblotting. Official HGNC symbols for the proteins in the figures are: RPT1 (also known as PSMC2), RPT2 (PSMC1), RPT3 (PSMC4), RPT4 (PSMC6), RPT5 (PSMC3), RPT6 (PSMC5), RPN1 (PSMD2), RPN2 (PSMD1), α7 (PSMA3) and RPN10 (PSMD4).

unexpectedly implies a more general conservation, that the subunit compositions of yeast and mammalian proteasomes are actually identical. While this paper was under review, the participation of Hsm3 and S5b in RP assembly was reported, consistent with our findings²⁶.

Nas6, Rpn14 and Hsm3 fit the definition of a molecular chaperone closely²⁷. They assist in assembly of an oligomeric complex, do not act as templates for the complex, and are absent from the mature complex. Chaperones typically function to prevent undesired and non-specific interactions of their target proteins, but RP chaperones represent an interesting case in which the interaction suppressed by the chaperone is a proper interaction in the mature complex, and moreover an interaction that is specifically required for assembly itself. These properties of RP chaperones point to a role in controlling the timing and ordering of maturation events.

How does competition of RP chaperones with the CP for occupancy of the Rpt C-domains promote proper base or Rpt ring assembly? In the simplest version of templated assembly, the ring would assemble in contact with the CP in a growing arc, subunit by subunit, until, with the inclusion of its last subunit, the ring is closed. In this ‘one-template model’, the Rpt–CP interaction, in combination with Rpt–Rpt interactions, would guide the proper choice of subunit and stabilize the nascent, open ring. Such a model was proposed for assembly of the CP β ring²⁸. Interestingly, β ring assembly is, like Rpt ring assembly, guided by the α ring, albeit the opposite face of the α ring. In the case of the Rpt ring, the one-template model may not apply, based in part on our identification of an apparent precursor to the base², BP1. This complex contains Rpn1, Rpt1, Rpt2, Rpt5 and Hsm3. Thus, the ring is likely to be assembled in patchwork fashion, partly from preformed ring segments.

The scaffold for BP1 appears to be a non-Rpt subunit, Rpn1 (ref. 2). Rpn2 is related to Rpn1, and may have a complementary scaffolding role for most or all base components that are absent from BP1. In this three-scaffold model, the CP α ring, Rpn1 and Rpn2 each help to template the ring. Both Rpn1 and Rpn2 are toroidal^{29,30}, a property likely to underlie their ability to nucleate ring assembly. A key feature of the three-template model is that the ordering of Rpt addition is achieved by interactions among multisubunit base precursors. An example of how RP chaperones could order the assembly pathway is that Hsm3 might help BP1 to assemble with other base precursors late in the assembly pathway by suppressing premature interaction of Rpt1 with the CP.

Rpt C termini prevent chaperone expulsion, even when the Rpt C terminus is productively docked in the CP pocket, indicating that RP chaperone expulsion is precise and requires a close approach of the CP to the RP². Despite the explanatory potential of this model, we do not exclude the possibilities that the RP chaperones may promote assembly through additional mechanisms, and perhaps regulate mature proteasomes in additional ways.

The ability of the all chaperones to bind Rpt C termini, compete with CP and promote assembly may reflect convergent evolution, considering that the chaperones are unlikely to have diverged from a common ancestor. For example, Nas6 is composed mainly of ankyrin repeats, whereas Hsm3 is rich in ARM repeats and Rpn14 in WD40 repeats. Key features of base assembly, including the BP1 intermediate, are conserved in mammals. The chaperone functions of yeast proteins Rpn14, Hsm3 and Nas6 are apparently paralleled in mammals by PAAF1, and S5b, and most interestingly the oncoprotein gankyrin. It is likely that gankyrin and S5b are misassigned as proteasome subunits. Thus, conservation of the assembly pathway

METHODS SUMMARY

Yeast strains, plasmids and antibodies. See Supplementary Table 2 for genotypes of strains used. Standard methods were used for strain construction and growth. The reported sequence of Hsm3 in the *Saccharomyces* genome database differs from the sequence in our strain background at the C terminus (see Supplementary Information). Plasmids and antibodies used are described in Supplementary Tables 3 and 4, respectively.

Immunoprecipitations and protein purifications. Recombinant proteins were expressed in and purified from Rosetta 2 cells (Novagen) as detailed in Supplementary Methods. For pulldown experiments, yeast cells were lysed by French press or glass beads. Lysate was cleared and incubated with the specific resin. For details, see Supplementary Information.

Native gels. For native gels, cell lysates were made in an ATP-containing buffer, cleared by centrifugation, and loaded on a 3.6% or 5.25% polyacrylamide gel. To transfer proteins from native gels to PVDF (polyvinylidene difluoride), the gels were pre-soaked for 10 min in 1× SDS–PAGE running buffer and transferred according to standard protocols. For two-dimensional immunoblotting, a lane was excised and loaded on top of an SDS–PAGE gel. Proteins were transferred from the second-dimension gel to PVDF and probed with the indicated antibodies. Detailed protocols for all experimental procedures can be found in Supplementary Information.

Received 24 October 2008; accepted 15 April 2009.

Published online 1 May 2009.

- Finley, D. Recognition and processing of ubiquitin–protein conjugates by the proteasome. *Annu. Rev. Biochem.* **78**, 477–513 (2009).
- Park, S. *et al.* Hexameric assembly of the proteasomal ATPases is templated through their C termini. *Nature* doi:10.1038/nature08065 (this issue).
- Leggett, D. S. *et al.* Multiple associated proteins regulate proteasome structure and function. *Mol. Cell* **10**, 495–507 (2002).
- Verma, R. *et al.* Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol. Biol. Cell* **11**, 3425–3439 (2000).
- Wang, X. *et al.* Mass spectrometric characterization of the affinity-purified human 26S proteasome complex. *Biochemistry* **46**, 3553–3565 (2007).
- Dawson, S., Higashitsuji, H., Wilkinson, A. J., Fujita, J. & Mayer, R. J. Gankyrin: a new oncoprotein and regulator of pRb and p53. *Trends Cell Biol.* **16**, 229–233 (2006).
- Guerrero, C., Milenkovic, T., Przulj, N., Kaiser, P. & Huang, L. Characterization of the proteasome interaction network using a QTAX-based tag-team strategy and protein interaction network analysis. *Proc. Natl Acad. Sci. USA* **105**, 13333–13338 (2008).
- Park, Y. *et al.* Proteasomal ATPase-associated factor 1 negatively regulates proteasome activity by interacting with proteasomal ATPases. *Mol. Cell. Biol.* **25**, 3842–3853 (2005).
- Lassot, I. *et al.* The proteasome regulates HIV-1 transcription by both proteolytic and nonproteolytic mechanisms. *Mol. Cell* **25**, 369–383 (2007).
- Collins, G. A. & Tansey, W. P. The proteasome: a utility tool for transcription? *Curr. Opin. Genet. Dev.* **16**, 197–202 (2006).
- Imai, J., Maruya, M., Yashiroda, H., Yahara, I. & Tanaka, K. The molecular chaperone Hsp90 plays a role in the assembly and maintenance of the 26S proteasome. *EMBO J.* **22**, 3557–3567 (2003).
- Isono, E. *et al.* The assembly pathway of the 19S regulatory particle of the yeast 26S proteasome. *Mol. Biol. Cell* **18**, 569–580 (2007).
- Kusmierczyk, A. R., Kunjappu, M. J., Funakoshi, M. & Hochstrasser, M. A multimeric assembly factor controls the formation of alternative 20S proteasomes. *Nature Struct. Mol. Biol.* **15**, 237–244 (2008).
- Kusmierczyk, A. R. & Hochstrasser, M. Some assembly required: dedicated chaperones in eukaryotic proteasome biogenesis. *Biol. Chem.* **389**, 1143–1151 (2008).
- Xie, Y. & Varshavsky, A. RPN4 is a ligand, substrate, and transcriptional regulator of the 26S proteasome: a negative feedback circuit. *Proc. Natl Acad. Sci. USA* **98**, 3056–3061 (2001).
- Kleijnen, M. F. *et al.* Stability of the proteasome can be regulated allosterically through engagement of its proteolytic active sites. *Nature Struct. Mol. Biol.* **14**, 1180–1188 (2007).
- Forster, A., Masters, E. I., Whitby, F. G., Robinson, H. & Hill, C. P. The 1.9 Å structure of a proteasome–11S activator complex and implications for proteasome–PAN/PA700 interactions. *Mol. Cell* **18**, 589–599 (2005).
- Gillette, T. G., Kumar, B., Thompson, D., Slaughter, C. A. & Demartino, G. N. Differential roles of the C-termini of AAA subunits of PA700 (19S regulator) in asymmetric assembly and activation of the 26S proteasome. *J. Biol. Chem.* **283**, 31813–31822 (2008).
- Smith, D. M. *et al.* Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's alpha ring opens the gate for substrate entry. *Mol. Cell* **27**, 731–744 (2007).
- Ammelburg, M., Frickey, T. & Lupas, A. N. Classification of AAA+ proteins. *J. Struct. Biol.* **156**, 2–11 (2006).
- Nakamura, Y. *et al.* Structural basis for the recognition between the regulatory particles Nas6 and Rpt3 of the yeast 26S proteasome. *Biochem. Biophys. Res. Commun.* **359**, 503–509 (2007).
- Nakamura, Y. *et al.* Structure of the oncoprotein gankyrin in complex with S6 ATPase of the 26S proteasome. *Structure* **15**, 179–189 (2007).
- Zhang, F. *et al.* Structural insights into the regulatory particle of the proteasome from *Methanocaldococcus jannaschii*. *Mol. Cell.* (in the press).
- Groll, M. *et al.* Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* **386**, 463–471 (1997).
- Gorbea, C., Taillandier, D. & Rechsteiner, M. Mapping subunit contacts in the regulatory complex of the 26S proteasome. S2 and S5b form a tetramer with ATPase subunits S4 and S7. *J. Biol. Chem.* **275**, 875–882 (2000).
- Le Tallec, B., Barrault, M. B., Guerois, R., Carre, T. & Peyroche, A. Hsm3/S5b participates in the assembly pathway of the 19S regulatory particle of the proteasome. *Mol. Cell* **33**, 389–399 (2009).
- Ellis, R. J. Molecular chaperones: assisting assembly in addition to folding. *Trends Biochem. Sci.* **31**, 395–401 (2006).
- Hirano, Y. *et al.* Dissecting beta-ring assembly pathway of the mammalian 20S proteasome. *EMBO J.* **27**, 2204–2213 (2008).
- Effantin, G., Rosenzweig, R., Glickman, M. & Steven, A. C. Electron microscopic evidence in support of α -solenoid models of proteasomal subunits Rpn1 and Rpn2. *J. Mol. Biol.* **386**, 1204–1211 (2009).
- Rosenzweig, R., Osmulski, P. A., Gaczynska, M. & Glickman, M. H. The central unit within the 19S regulatory particle of the proteasome. *Nature Struct. Mol. Biol.* **15**, 573–580 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Mann for Rpt1 (also known as Cim5) and Rpt6 (Cim3) antibodies, K. Kubota for preparing HeLa lysates and especially L. Huang for the human Rpn11–HTBH-expressing cell line. We also thank J. Hanna, M. Schmidt and members of the Finley laboratory, in particular S. Elsasser, for critically reading the manuscript. This work was supported by the NIH (GM043601 to D.F. and GM67945 to S.P.G.), a NIH NRSA postdoctoral fellowship (5F32GM75737-2 to S.P.) and an EMBO long-term fellowship (to J.R.).

Author Contributions J.R., Y.H. and F.E.M. performed the experiments. W.H., F.E.M. and S.P.G. performed mass spectrometry. G.T. performed modelling. F.Z. and Y.S. contributed to modelling. B.-H.L. purified human proteasomes. J.R., S.P. and D.F. interpreted data and developed the model. J.R. and D.F. planned the studies and wrote the manuscript. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.F. (daniel_finley@hms.harvard.edu).

LETTERS

Hexameric assembly of the proteasomal ATPases is templated through their C termini

Soyeon Park¹, Jeroen Roelofs¹, Woong Kim¹, Jessica Robert¹, Marion Schmidt², Steven P. Gygi¹ & Daniel Finley¹

Substrates of the proteasome are recognized and unfolded by the regulatory particle, and then translocated into the core particle (CP) to be degraded¹. A hetero-hexameric ATPase ring, containing subunits Rpt1–6, is situated within the base subassembly of the regulatory particle¹. The ATPase ring sits atop the CP, with the Rpt carboxy termini inserted into pockets in the CP^{2–6}. Here we identify a previously unknown function of the Rpt proteins in proteasome biogenesis through deleting the C-terminal residue from each Rpt in the yeast *Saccharomyces cerevisiae*. Our results indicate that assembly of the hexameric ATPase ring is templated on the CP. We have also identified an apparent intermediate in base assembly, BP1, which contains Rpn1, three Rpts and Hsm3, a chaperone for base assembly. The Rpt proteins with the strongest assembly phenotypes, Rpt4 and Rpt6, were absent from BP1. We propose that Rpt4 and Rpt6 form a nucleating complex to initiate base assembly, and that this complex is subsequently joined by BP1 to complete the Rpt ring. Our studies show that assembly of the proteasome base is a rapid yet highly orchestrated process.

The core particle (20S CP) of the 26S proteasome is associated with the regulatory particle (19S RP or PA700), which in turn consists of two subassemblies, the base and lid¹. The CP is composed of α -type and β -type subunits, with each family of subunits arranged into hetero-heptameric rings, which are stacked to form a barrel-like $\alpha_7\beta_7\beta_7\alpha_7$ complex^{1,7}. The proteolytic sites are inside the CP barrel, and substrate access to these sites is achieved by opening a gate in the α ring. Binding of the RP opens the gate. Rpt1–6, six ATPases of the base, are thought to form a hetero-hexameric ring, and the C termini of these Rpts insert into pockets in the CP α ring^{2–6}. The extreme C termini of four Rpts contain the motif Hb-Y-X (hydrophobic residue-Tyr-X, where X denotes any amino acid), which mediates opening of the CP gate^{3,4,6} (Supplementary Table 1).

The C-terminal carboxylate of a given Rpt protein forms a salt bridge to a lysine residue at the bottom of the CP α -ring pocket^{2,3}. To perturb these salt bridges, we deleted the C-terminal residue from each Rpt protein, generating six chromosomal mutations (*rpt1*- Δ 1 to *rpt6*- Δ 1). Proteasomes in whole-cell extracts from each mutant were analysed by native polyacrylamide gel electrophoresis (PAGE), combined with an in-gel peptidase assay using a fluorogenic substrate, LLVY-AMC (Fig. 1a). Free CP was strongly increased in *rpt3*- Δ 1, *rpt4*- Δ 1, *rpt5*- Δ 1 and *rpt6*- Δ 1 mutants, as revealed by the addition of 0.02% SDS (Fig. 1a). As previously shown, SDS opens the CP gate, which is closed in the free CP^{1,7}. The CP, and a representative CP subunit, α_7 , were most strongly induced in *rpt4*- Δ 1 and *rpt6*- Δ 1 mutants (Fig. 1a, b). Interestingly, Rpt4 and Rpt6 are the two Rpts that lack the Hb-Y-X gating motifs⁴ (Supplementary Table 1). These mutants also showed strong reductions in proteasome holoenzyme (Fig. 1a, RP-CP and RP₂-CP; these are alternative forms of holoenzyme containing one or two RP complexes).

CP induction was mediated principally by Rpn4, a transcription factor that upregulates the expression of proteasome subunits when proteasome activity is compromised^{8,9}; *RPN4* deletion in *rpt4*- Δ 1 or *rpt6*- Δ 1 mutants markedly decreased the level of α_7 and free CP (Fig. 1c and Supplementary Fig. 1). Assembly defects in the *rpt* mutant proteasomes correlated with the accumulation of ubiquitin conjugates and, in the case of the strongest mutant, *rpt6*- Δ 1, a growth defect at 30 °C (Fig. 1b, d). Phenotypes of *rpt4*- Δ 1 and *rpt6*- Δ 1 were exacerbated by *RPN4* deletion, suggesting that the proteasome stress response masks the severity of these mutations (Fig. 1e).

Proteolytically inactive subassemblies such as the RP and lid cannot be visualized by activity assays. Therefore, to address the assembly defects in *rpt4*- Δ 1 and *rpt6*- Δ 1 mutants further, proteasomes were analysed by immunoblotting after native PAGE. Given the apparent role of Rpt C termini in CP binding^{2–6}, the increase in free CP and decrease in RP₂-CP and RP-CP might predict an increase in free RP in *rpt4*- Δ 1 and *rpt6*- Δ 1 mutants. In contrast, free RP was not detected in these mutants; instead, free lid was observed (Fig. 1f and Supplementary Fig. 2). The production of free lid was unexpected because the Rpt C termini, if docking to the CP, cannot be in contact with or proximal to the lid. The base-CP complex, previously observed in a lid mutant¹⁰, was not apparent in *rpt4*- Δ 1 or *rpt6*- Δ 1 mutants (Fig. 1f and Supplementary Fig. 2). Thus, productive interaction between the CP and the Rpt C termini of the base is apparently required for stable incorporation of the lid to complete proteasome assembly.

Because the base joins the lid to the CP in proteasome holoenzymes, the accumulation of free lid and free CP in *rpt4*- Δ 1 and *rpt6*- Δ 1 mutants may reflect underlying defects in base assembly^{11,12}. The base can be produced by exposing the assembled proteasome to 1 M NaCl^{13,14}. The resulting particle represents the mature form of the base (Supplementary Fig. 3). If the base exists as a free endogenous complex, its isolation should be possible by affinity purification via protein-A-tagged Rpt1, using a physiological salt concentration. Such samples were analysed by two-dimensional native SDS-PAGE. A base variant, which we refer to as base*, was detected by this method (Fig. 2a, top). Interestingly, mass spectrometric analysis revealed that base* contains, in addition to Ubp6 and subunits of the base¹⁵, three proteins that are not proteasome components: Rpn14, Nas6 and Hsm3 (Fig. 2a). An accompanying paper¹⁶ shows that these three proteins function as molecular chaperones and more specifically as base assembly factors¹⁶. We show here that the function of these proteins, termed RP chaperones, is tightly linked to the Rpt C termini.

In *rpt4*- Δ 1 mutants, base* levels are severely reduced, suggesting a defect in base assembly (Fig. 2a, b). Base* can also be visualized in unfractionated extract from untagged wild-type cells (Fig. 2c, d). Base* was deficient in *rpt6*- Δ 1 cells (Fig. 2c) and undetectable in the RP-chaperone triple mutant *hsm3* Δ *nas6* Δ *rpn14* Δ (Fig. 2d).

¹Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA. ²Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Forchheimer Building, Room 305, Bronx, New York 10461, USA.

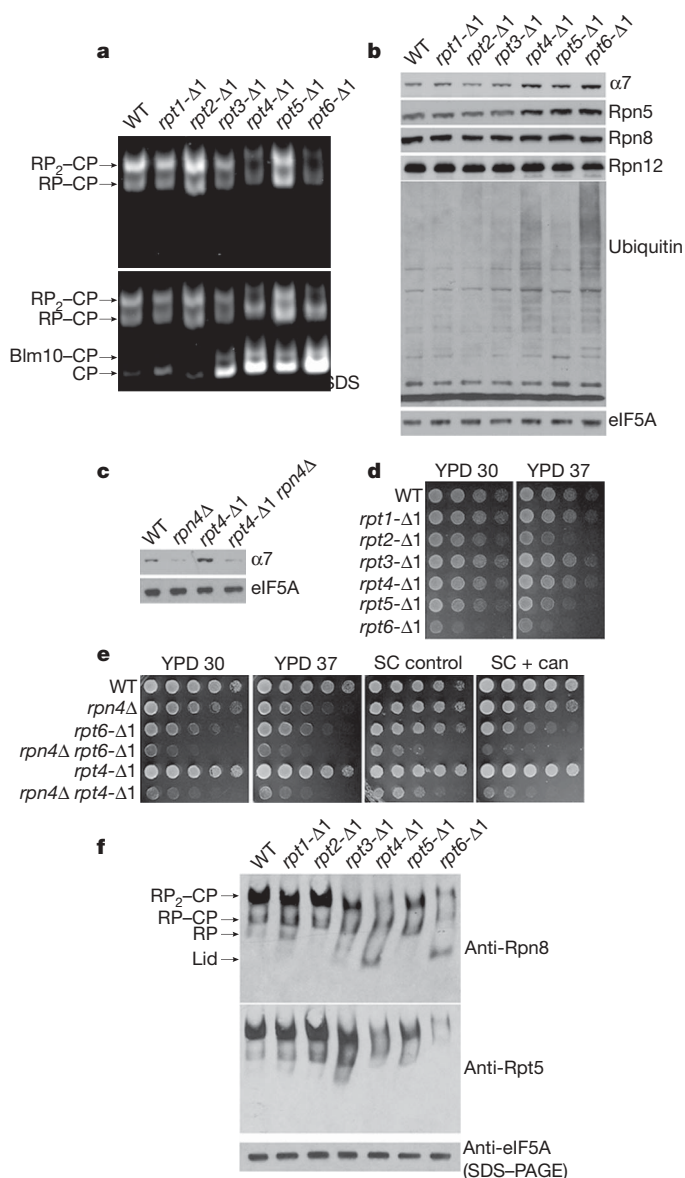


Figure 1 | *rpt4-Δ1* and *rpt6-Δ1* mutants are proteasome hypomorphs with defective proteasome assembly. **a**, Native PAGE (3.5%) and two consecutive LLVY-AMC assays with whole-cell extracts (85 μg). Immediately after the first LLVY-AMC assay (top panel), the second assay was conducted in the presence of 0.02% SDS (bottom panel). Blm10 is a protein that associates with the cylinder end of the CP. **b**, **c**, SDS-PAGE and immunoblotting with whole-cell extracts. eIF5A is a loading control. **d**, **e**, Growth phenotypes of *rpt* mutants. Cells were spotted in fourfold dilutions on rich medium (YPD²⁴), synthetic medium (SC, control) or synthetic medium containing canavanine (can) at 0.5 μg ml⁻¹. Plates were incubated for 2–4 days at 30 °C (YPD 30) or at 37 °C (YPD 37). **f**, 3.5% native PAGE and immunoblotting of whole-cell extracts (85 μg) against Rpn8 (lid subunit) and Rpt5 (base subunit). Extracts (20 μg) were resolved by SDS-PAGE and immunoblotting for loading control.

The disappearance of base* in *rpt4-Δ1* mutants suggested an upstream assembly defect that would be manifest in the accumulation of smaller assembly intermediates of the base. Accordingly, affinity purification from protein A (*ProA*)-*TeV*-*Rpt1* *rpt4-Δ1* mutants revealed a previously unknown species that we termed base precursor 1, or BP1 (Fig. 2a, b; the *TeV* element is a protease cleavage site). BP1 has two forms. The lower-mobility form contains deubiquitinating enzyme Ubp6, and is referred to as Ubp6 + BP1 (Fig. 2a, b). BP1 was also observed in wild type, although at lower levels (Fig. 2a, b). The proportion of BP1 loaded with Ubp6 is higher in the mutant; it is likely that base assembly, which consumes BP1, is faster than Ubp6 loading

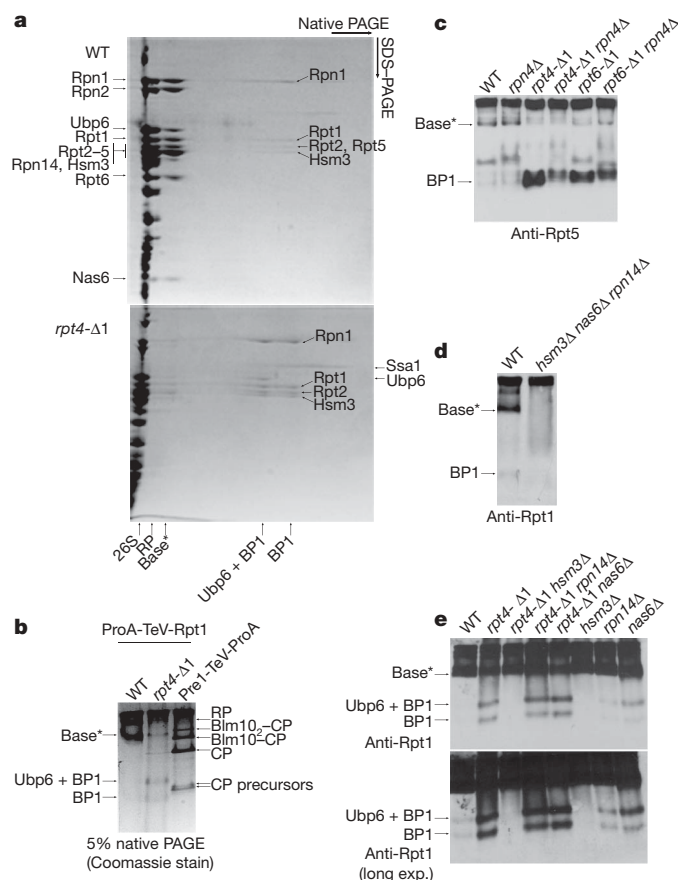


Figure 2 | Identification of a base assembly intermediate. **a**, Two-dimensional native SDS-PAGE of affinity purifications with ProA-*TeV*-*Rpt1* (protein A tag appended to the N terminus of *Rpt1*, a base subunit). After first-dimension native PAGE (5%), native gel lanes were individually excised and subjected to second-dimension SDS-PAGE (12.5%). Gels were stained with Coomassie blue. Individual spots of the base*, BP1 and Ubp6 + BP1 complexes were excised for mass spectrometry. Labels on the left indicate spots within base*. The presence of Rpn10 and Rpn13 in base* was not determined. Note that Rpt5 is absent from BP1 purified from *rpt4-Δ1* mutants, but present in BP1 in untagged *rpt4-Δ1* extracts (**c**). This seems to reflect a labile association of Rpt5 with BP1 in *rpt4-Δ1* mutants. The presence of Ubp6 in BP1 is explained by its interaction with Rpn1 (ref. 26). Note that base* and BP1 species appear comparable between *rpt6-Δ1* and *rpt4-Δ1* mutants in their level and composition. **b**, 5% native PAGE following affinity purification using a ProA-*TeV*-tagged *Rpt1* as in **a** or a *TeV*-ProA-tagged Pre1 (CP subunit). Native gels were stained with Coomassie blue. **c**, **d**, Immunoblotting following 5% native PAGE of whole-cell extracts (100 μg). **e**, Immunoblotting following 5% native PAGE of affinity-purified samples (2 μg) from indicated strains carrying ProA-*TeV*-*Rpt1*.

in wild type, but slower than Ubp6 loading in the mutant (see below). By mass spectrometry after two-dimensional native SDS-PAGE, BP1 was found to contain Rpn1, Rpt1, Rpt2, Rpt5 and the RP chaperone Hsm3 (Fig. 2a). The presence of Hsm3 in BP1 suggests that BP1 is an assembly intermediate. BP1 was detected in untagged whole-cell extracts, and thus, like base*, it is not an artefact of tagging or purification (Fig. 2c, d). BP1 levels in extracts are increased in *rpt4-Δ1* and *rpt6-Δ1* mutants (Fig. 2c). In contrast, *HSM3* deletion resulted in an apparently complete elimination of BP1 (Fig. 2e). Thus, Hsm3 is not a bystander in BP1, but strongly regulates its abundance.

The absence of RP chaperones Rpn14 and Nas6 from BP1 can be explained by their binding specificities; whereas Hsm3 binds to the BP1 component Rpt1 (refs 12, 16), the other RP chaperones bind to Rpt proteins that are absent from BP1: Nas6 to Rpt3 (refs 16, 17) and Rpn14 to Rpt4 and Rpt6 (ref. 16). RP chaperones bind to the C-terminal domains (C-domains) of Rpt proteins, but not to their extreme C termini¹⁶. In contrast to *HSM3*, deletion of *RPN14* or

NAS6 had little effect on the level of BP1 (Fig. 2e), indicating that their roles in base assembly are distinct from the regulation of BP1.

The absence of Rpt4 and Rpt6 from BP1 is interesting because these are the Rpt proteins whose C termini are most critical for base assembly. The accumulation of BP1 in *rpt4*- Δ 1 and *rpt6*- Δ 1 mutants probably reflects perturbation of an assembly event upstream of BP1 incorporation into nascent base. Also absent from BP1 were CP subunits (Supplementary Fig. 4); the failure to observe a stable BP1–CP complex suggests that such complexes are not involved in assembly. Instead, BP1 may load onto a separate, complementary assembly intermediate that includes Rpt4 and Rpt6 and is capable of direct interaction with CP.

To verify that BP1 is a nascent complex used for proteasome assembly, we performed a pulse-chase experiment using *in vivo* S^{35} metabolic labelling (Fig. 3). BP1 was visualized by pulse-labelling, and, after a 30-min chase with excess unlabelled methionine, S^{35} -labelled BP1 was substantially decreased (Fig. 3). Thus, BP1 is a transient complex *in vivo*. Consistent with steady-state experiments, S^{35} -labelled BP1 was more prominent in *rpt4*- Δ 1 than in wild type, suggesting that its incorporation into nascent base depends on proper interaction between Rpt4 C termini and CP.

The interaction of CP with Rpt4 and Rpt6 is regulated by Rpn14, which specifically interacts with the C-domains of these two Rpts¹⁶. Rpn14 associates with RP or base, but little or no Rpn14 is found on CP-containing complexes such as the proteasome holoenzyme¹⁶. Notably, Rpn14 was readily detected on Rpt6- Δ 1 proteasome holoenzyme that had been affinity-purified via a CP tag (Fig. 4a and Supplementary Fig. 5). In contrast, Hsm3 and Nas6 did not

associate with Rpt6- Δ 1 proteasomes (Fig. 4a), showing that the effect of the *rpt6*- Δ 1 mutation was local and restricted to the binding partner of Rpt6. We conclude that Rpn14 is released as a specific consequence of the engagement of the C terminus of its cognate Rpt with the CP. Proteasome holoenzyme or base–CP assembly *per se* is not sufficient for displacement. The RP chaperone competes with CP for occupancy of the Rpt C-domain, providing a simple mechanism by which RP chaperones can regulate base assembly and subsequently exit the complex¹⁶. It is possible that the failure to release Rpn14 could account for the phenotypic defects of *rpt4*- Δ 1 and *rpt6*- Δ 1 mutants; however, this is unlikely, because *rpn14* Δ is not a suppressor of either allele (data not shown).

Analysis of Rpt3- Δ 1 proteasomes also revealed a defect in chaperone expulsion during assembly. Instead of Rpn14, Rpt3- Δ 1 proteasomes retained Nas6, the binding partner of Rpt3 (refs 16 and 18, and Fig. 4b). The *rpt3*- Δ 1 mutation had no effect on the association of Hsm3 or Rpn14 with proteasomes. The association of Rpn14 or Nas6 with the holoenzyme was not due to an increase in their protein expression level in *rpt6*- Δ 1 or *rpt3*- Δ 1 mutants, respectively (data not shown).

Our structural modelling suggested that competition between CP and RP chaperones for Rpt binding reflects the proximity of their binding sites¹⁶. The model predicts that separating the two binding

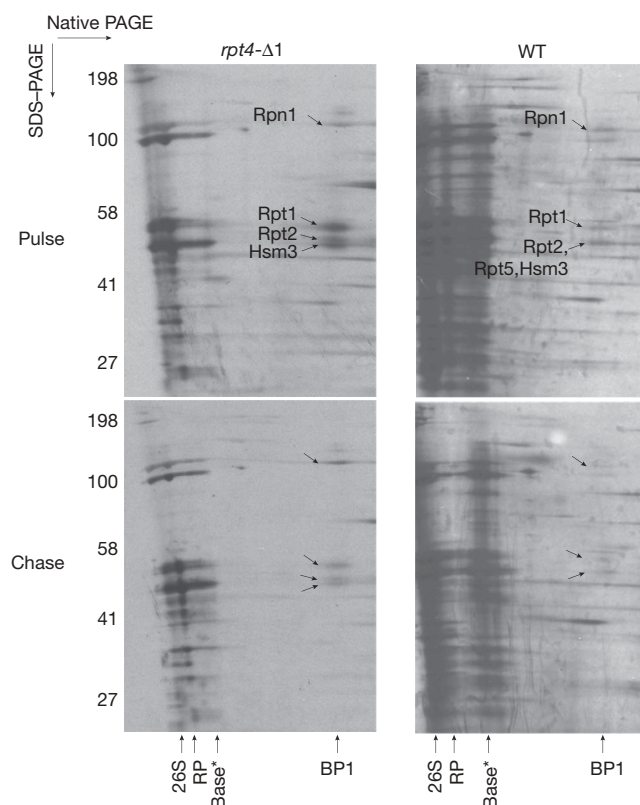


Figure 3 | Pulse-chase analysis of the BP1 base assembly intermediate. Strains were metabolically labelled with S^{35} -methionine *in vivo* for 4–5 min (pulse) and chased with excess methionine for 30 min (chase). Samples were then subjected to affinity purification with ProA–TeV–Rpt1 and two-dimensional native SDS–PAGE (5% and 4–12%, respectively) followed by autoradiography. Molecular masses are indicated at left in kDa. BP1, but not Ubp6 + BP1, was labelled with S^{35} -methionine during the pulse in both wild type and the *rpt4*- Δ 1 mutant.

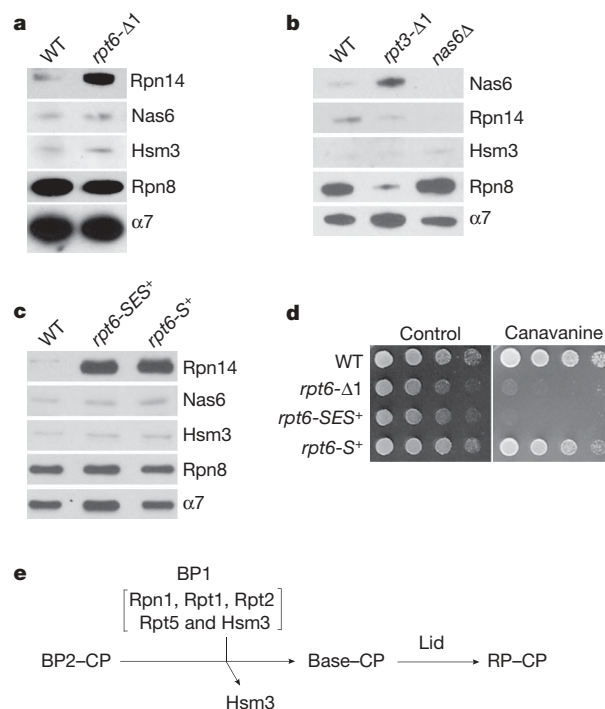


Figure 4 | Interactions between Rpt C termini and the CP control RP-chaperone release. **a–c**, SDS–PAGE and immunoblotting followed by affinity purification using a CP tag (Pre1–TeV–ProA). Rpn8 (lid subunit) and α 7 (CP subunit) are loading controls. The Nas6 Δ proteasome (**b**) is a negative control. The *rpt6*-SES⁺ or *rpt6*-S⁺ mutants in **c** contain either a three (Ser–Glu–Ser) or one (Ser) residue insertion before the fourth residue from the C terminus. **d**, Phenotypic analysis of RPT6 insertion mutants. Fourfold serial dilutions were spotted on synthetic medium (control) or canavanine (1 μ g ml^{–1}) plates. Cells were grown for 2–4 days at 30 °C. **e**, Model for late-stage base assembly. It is proposed that the base is formed by the addition of the BP1 complex, containing Rpn1, Rpt1, Rpt2, Rpt5 and Hsm3, to the putative BP2 complex. Initial contact with the CP is provided by BP2, whereas BP1 exists independently of the CP until it joins the BP2–CP complex. BP2 has not been identified and its exact composition is unknown. However, probable components of BP2 are Rpt4, Rpt6, Rpn2 and perhaps Rpt3. Rpn14 and Nas6 (not shown) may function before the formation of BP2–CP.

sites would abrogate competition and thus allow retention of RP chaperones on the proteasome. Remarkably, insertion of a single residue into Rpt6 (*rpt6-S⁺*) four residues from the C terminus prevented expulsion of Rpn14 (Fig. 4c). Thus, the expulsion mechanism is very precise. Importantly, the failure to expel Rpn14 was not a result of failure to dock the mutant Rpt6 C terminus at the CP, because *rpt6-S⁺* mutants did not phenocopy *rpt6-Δ1* mutants (Fig. 4d). The release of non-cognate RP chaperones was unaffected by C-terminal insertions (Fig. 4c).

In bacterial AAA proteins, formation of the ATPase ring is thought to be spontaneous and unassisted¹⁹. In contrast, the proteasome's ATPase ring is assembled through a programmatic and remarkably intricate pathway. This and the accompanying study¹⁶ suggest that the pathway of base assembly involves four distinct classes of factors, each found in multiple forms: the seven-membered α -subunit ring of the CP, the six ATPases of the Rpt ring, two scaffolding factors (Rpn1 and Rpn2) and three RP chaperones.

The importance for assembly of the extreme C termini of the Rpts indicates a prominent role of CP templating in Rpt ring formation. We propose that one subset of Rpts assembles directly on the CP, including Rpt4 and Rpt6, the mutants of which show the strongest assembly phenotypes. In contrast, another group of Rpts, including Rpt1, Rpt2 and Rpt5, is proposed to initially assemble free of the CP in the BP1 complex (Fig. 4e). Because BP1 disappears rapidly in pulse-chase experiments, it behaves as a true *in vivo* assembly intermediate. The interaction of the Rpt binding proteins with the α pockets of the CP appears to be negatively regulated by the RP chaperones¹⁶. These chaperones may suppress α -pocket–Rpt interactions during certain stages of assembly, and, during other stages, disengage from the Rpt to permit assembly to precede forward (Fig. 4). In this way the RP chaperones may temporally sequence the assembly process.

A simple although hypothetical scenario is that the base is formed by addition of BP1 to a preformed complex containing some or all of the base components absent from BP1. This proposed assembly intermediate is termed BP2 in our three-template model (Fig. 4e). Rpt4 and Rpt6 are proposed to mediate BP2 formation by CP templating. Slow maturation of BP2 may explain the accumulation of BP1 observed in *rpt4-Δ1* and *rpt6-Δ1* mutants, consistent with this model. We also propose that Rpn2 functions as a scaffold to organize BP2, on the basis of evidence found in this and previous studies. For example, Rpt4 (ref. 20) and Rpt6 (ref. 21) have been shown to specifically interact with Rpn2 (whereas Rpt1 and Rpt2 interact with the Rpn1 (ref. 22) scaffold of BP1). Moreover, Rpn2 binds CP directly²³. Thus, a complex containing Rpn2, Rpt4 and Rpt6 is well suited to nucleate base assembly. Rpn2 and the CP may jointly template early ring assembly as the CP–Rpt binding events are probably not strong enough to single-handedly specify Rpt ring assembly. A base subassembly containing Rpn1 and Rpn2 was isolated by *in vitro* dissociation of proteasomes²³, but according to our findings it might not be formed during proteasome biogenesis. However, the Rpn1–Rpn2 interaction could be important for driving association of the BP1 complex with a complementary Rpn2-containing complex, such as the putative BP2.

METHODS SUMMARY

Purification of BP1. Strains carrying ProA–TeV–Rpt1 were grown in 6 l of minimal (SD²⁴) medium to D_{600} of 0.8–1.2. Cells were collected, washed once with ice-cold water and drop-frozen in liquid nitrogen. Frozen yeast samples were then ground using an MM301 grinding mill (Restch) under liquid nitrogen following the manufacturer's instructions, or using a mortar and pestle as previously described²⁵. Ground powder was hydrated in proteasome buffer (50 mM Tris–HCl, pH 7.5, 5 mM MgCl₂, 1 mM EDTA and 10% glycerol) supplemented with 2 mM ATP, protease inhibitor tablets (Complete, Roche), 2 mM PMSF, 1 mM benzamidine, 10 μ g ml^{−1} pepstatin A and 1 μ g ml^{−1} antipain. Cell extracts were cleared at 30,000g for 30 min at 4 °C, and the supernatants were mixed with rabbit immunoglobulin G resin (Cappel, MP Biomedicals) for 90 min at 4 °C. Resins were collected at 900g for 2 min at 4 °C and washed with proteasome

buffer containing 50 mM NaCl three times, followed by a final wash with proteasome buffer alone. BP1 was then released from the resin by incubating with recombinant AcTEV protease (Invitrogen) at 2.5 units l^{−1} culture in proteasome buffer containing 2 mM ATP and protease inhibitors for 1 h at 30 °C. Eluates were concentrated using Ultrafree-0.5 centrifugal filter device with 30 kDa NMWL (Millipore).

Pulse-chase analysis of BP1, native PAGE and immunoblot. Detailed methods and strain list are described in Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 October 2008; accepted 9 April 2009.

Published online 1 May 2009.

1. Finley, D. Recognition and processing of ubiquitin–protein conjugates by the proteasome. *Annu. Rev. Biochem.* **78**, 477–513 (2009).
2. Forster, A., Masters, E. I., Whitby, F. G., Robinson, H. & Hill, C. P. The 1.9 Å structure of a proteasome–11S activator complex and implications for proteasome–PAN/PA700 interactions. *Mol. Cell* **18**, 589–599 (2005).
3. Rabl, J. *et al.* Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. *Mol. Cell* **30**, 360–368 (2008).
4. Smith, D. M. *et al.* Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's alpha ring opens the gate for substrate entry. *Mol. Cell* **27**, 731–744 (2007).
5. Whitby, F. G. *et al.* Structural basis for the activation of 20S proteasomes by 11S regulators. *Nature* **408**, 115–120 (2000).
6. Gillette, T. G., Kumar, B., Thompson, D., Slaughter, C. A. & Demartino, G. N. Differential roles of the C-termini of AAA subunits of PA700 (19S regulator) in asymmetric assembly and activation of the 26S proteasome. *J. Biol. Chem.* **283**, 31813–31822 (2008).
7. Borissenko, L. & Groll, M. 20S proteasome and its inhibitors: crystallographic knowledge for drug development. *Chem. Rev.* **107**, 687–717 (2007).
8. Mannhaupt, G., Schnall, R., Karpov, V., Vetter, I. & Feldmann, H. Rpn4p acts as a transcription factor by binding to PACE, a nonamer box found upstream of 26S proteasomal and other genes in yeast. *FEBS Lett.* **450**, 27–34 (1999).
9. Xie, Y. & Varshavsky, A. RPN4 is a ligand, substrate, and transcriptional regulator of the 26S proteasome: a negative feedback circuit. *Proc. Natl Acad. Sci. USA* **98**, 3056–3061 (2001).
10. Isono, E. *et al.* The assembly pathway of the 19S regulatory particle of the yeast 26S proteasome. *Mol. Biol. Cell* **18**, 569–580 (2007).
11. Kusmierczyk, A. R., Kunjappu, M. J., Funakoshi, M. & Hochstrasser, M. A multimeric assembly factor controls the formation of alternative 20S proteasomes. *Nature Struct. Mol. Biol.* **15**, 237–244 (2008).
12. Le Tallec, B., Barrault, M. B., Guerois, R., Carre, T. & Peyroche, A. Hsm3/S5b participates in the assembly pathway of the 19S regulatory particle of the proteasome. *Mol. Cell* **33**, 389–399 (2009).
13. Saeki, Y., Toh-e, A. & Yokosawa, H. Rapid isolation and characterization of the yeast proteasome regulatory complex. *Biochem. Biophys. Res. Commun.* **273**, 509–515 (2000).
14. Leggett, D. S., Glickman, M. H. & Finley, D. Purification of proteasomes, proteasome subcomplexes, and proteasome-associated proteins from budding yeast. *Methods Mol. Biol.* **301**, 57–70 (2005).
15. Glickman, M. H. *et al.* A subcomplex of the proteasome regulatory particle required for ubiquitin–conjugate degradation and related to the COP9–signalosome and eIF3. *Cell* **94**, 615–623 (1998).
16. Roelofs, J. *et al.* Chaperone-mediated pathway of proteasome regulatory particle assembly. *Nature* doi:10.1038/nature08063 (this issue).
17. Nakamura, Y. *et al.* Structural basis for the recognition between the regulatory particles Nas6 and Rpt3 of the yeast 26S proteasome. *Biochem. Biophys. Res. Commun.* **359**, 503–509 (2007).
18. Dawson, S. *et al.* Gankyrin is an ankyrin-repeat oncoprotein that interacts with CDK4 kinase and the S6 ATPase of the 26 S proteasome. *J. Biol. Chem.* **277**, 10893–10902 (2002).
19. Kress, W., Mutschler, H. & Weber-Ban, E. Assembly pathway of an AAA+ protein: tracking ClpA and ClpAP complex formation in real time. *Biochemistry* **46**, 6183–6193 (2007).
20. Richmond, C., Gorbea, C. & Rechsteiner, M. Specific interactions between ATPase subunits of the 26 S protease. *J. Biol. Chem.* **272**, 13403–13411 (1997).
21. Hartmann-Petersen, R., Tanaka, K. & Hendil, K. B. Quaternary structure of the ATPase complex of human 26S proteasomes determined by chemical cross-linking. *Arch. Biochem. Biophys.* **386**, 89–94 (2001).
22. Gorbea, C., Taillandier, D. & Rechsteiner, M. Mapping subunit contacts in the regulatory complex of the 26S proteasome. S2 and S5b form a tetramer with ATPase subunits S4 and S7. *J. Biol. Chem.* **275**, 875–882 (2000).
23. Rosenzweig, R., Osmulski, P. A., Gaczynska, M. & Glickman, M. H. The central unit within the 19S regulatory particle of the proteasome. *Nature Struct. Mol. Biol.* **15**, 573–580 (2008).
24. Guthrie, C. & Fink, G. R. *Guide to Yeast Genetics and Molecular Biology*. (Academic Press, 1991).

25. Verma, R. *et al.* Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol. Biol. Cell* **11**, 3425–3439 (2000).
26. Leggett, D. S. *et al.* Multiple associated proteins regulate proteasome structure and function. *Mol. Cell* **10**, 495–507 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. D. Orth for discussion and technical assistance, and C. Hill for discussions. We also thank J. W. Hanna, S. Elsasser and all members

of Finley laboratory for comments on the manuscript. The work was supported by grants from US National Institutes of Health (NIH) to D.F. (GM043601) and S.P.G. (GM67945), an EMBO long-term fellowship to J.R. and an NIH NRSA postdoctoral fellowship (5F32GM75737-2) to S.P.

Author Contributions S.P. conducted all experiments. J.R. and D.F. contributed to designing experiments. W.K. performed mass spectrometry. S.P. and D.F. wrote the paper. All authors provided comments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.F. (daniel_finley@hms.harvard.edu).

METHODS

Yeast strains and media. All yeast manipulations were conducted according to standard protocols²⁴. Strains used in this study are listed in Supplementary Table 2. C-terminal single deletions in each of the six *RPT* genes were made in their endogenous chromosomal loci by integrative recombination essentially as described in a previous study⁴. Strains were grown in YPD medium to D_{600} of 0.8–1.2 for the analysis of BP1, or to D_{600} 2.0–4.0 for the analysis of mature proteasomes.

Native PAGE, in-gel peptidase assay, two-dimensional native SDS–PAGE and immunoblot. Proteasome holoenzymes were analysed by 3.5% native PAGE and in-gel peptidase assay as described previously¹⁴. BP1 and base complexes were resolved by 5% discontinuous native PAGE using conditions described in a previous study²⁷. For immunoblotting, native gels were incubated in 1× SDS–PAGE running buffer²⁸ (or transfer buffer²⁸ containing 0.1% SDS) for 15 min and transferred to PVDF at 80 mA overnight at 4 °C. The procedure for two-dimensional native SDS–PAGE is described in detail in a previous study²⁸. Rpn5, Rpn8, Rpn12 and eIF5A antibodies were previously described²⁹. Antibodies against Rpt5 and $\alpha 7$ were from Biomol and were used at 1:2,000. Rpt1 antibody was used as previously described²⁹. RP chaperone antibodies are described in the accompanying study¹⁶.

Purification of proteasomal subcomplexes. 26S proteasome holoenzyme, 19S RP, 20S CP and lid were affinity-purified using a ProA tag in the presence of 1 mM ATP and 5 mM $MgCl_2$ following previously described procedures¹⁴. For the affinity-purification of the base, cells were lysed under liquid nitrogen as described in the Methods Summary and were subjected to a previously described procedure¹⁴. Purified base resulting from liquid nitrogen lysis migrated on native gels.

The endogenous base (base*) and the BP1 complexes were affinity-purified as described in the Methods Summary. During purification, the beads were washed with 50 mM NaCl. Increasing the concentration of salt in the wash buffer to 150 mM NaCl released Ubp6 from the Ubp6 + BP1 complex, but did not disrupt the integrity of the base* or BP1 complexes (data not shown).

Mass spectrometry of BP1 subunits. BP1 subunit bands were Coomassie-stained after two-dimensional native SDS–PAGE. Each subunit band was excised and digested in-gel with trypsin. The resulting peptides were separated by nanoscale reversed-phase liquid chromatography coupled to a tandem mass spectrometer (LTQ and LTQ XL, Thermo Electron). Each mass spectrometry (MS) scan was followed by collecting ten MS/MS spectra from the most abundant ions³⁰.

MS/MS spectra were searched against a yeast database using the Sequest algorithm³¹.

Cell lysis. French press lysis was used for experiments shown in Fig. 1a, f and Supplementary Fig. 2. Liquid nitrogen lysis as described in Methods Summary was used for experiments shown in Figs 2, 3 and 4a–c and Supplementary Figs 1, 3, 4 and 5.

Pulse-chase analysis of BP1. Strains were grown in 1 l of minimal (SD²⁴) medium to a D_{600} of 0.6. Cells were then collected and resuspended in 1 l of SD-methionine medium for 1 h. Cells were pelleted and divided into two 25-ml samples in SD-methionine medium. Each sample was pulsed with 3.5 mCi of S³⁵-methionine using EasyTag EXPRESS S³⁵ protein labelling mix (Perkin Elmer) for 5 min (*PROA-TEV-RPT1 rpt4-Δ1*) or 4 min (*PROA-TEV-RPT1*), and labelled samples were collected by centrifugation at 3,000g for 2 min at 4 °C. For the pulsed sample, the labelled cells were drop-frozen in liquid nitrogen immediately. For the chased sample, the labelled cells were resuspended in 25 ml of s.d. medium with excess methionine freshly added to 0.5 mg ml^{−1} final concentration and further incubated at 30 °C for 30 min. At the end of the chase, cells were collected at 3,000g for 2 min and drop-frozen in liquid nitrogen. Frozen yeast cells were then ground under liquid nitrogen with a pestle in a mortar nestled in dry ice. BP1 was immunoprecipitated via the ProA tag and eluted with AcTEV protease as described in Methods Summary. The entire eluate was subjected to two-dimensional native SDS–PAGE (5% and 4–12%, respectively). Gels were dried at 80 °C for 1 h and exposed to X-ray film at −80 °C with intensifier screen for 4 days (*PROA-TEV-RPT1 rpt4-Δ1*) to 2 months (*PROA-TEV-RPT1*).

27. Li, X., Kusmierczyk, A. R., Wong, P., Emili, A. & Hochstrasser, M. Beta-subunit appendages promote 20S proteasome assembly by overcoming an Ump1-dependent checkpoint. *EMBO J.* **26**, 2339–2349 (2007).
28. Kleijnen, M. F. *et al.* Stability of the proteasome can be regulated allosterically through engagement of its proteolytic active sites. *Nature Struct. Mol. Biol.* **14**, 1180–1188 (2007).
29. Hanna, J. *et al.* Deubiquitinating enzyme Ubp6 functions noncatalytically to delay proteasomal degradation. *Cell* **127**, 99–111 (2006).
30. Peng, J. & Gygi, S. P. Proteomics: the move to mixtures. *J. Mass Spectrom.* **36**, 1083–1091 (2001).
31. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).

An unusual carbon–carbon bond cleavage reaction during phosphinothricin biosynthesis

Robert M. Cicchillo^{1,2,3*}, Houjin Zhang^{2,4*}, Joshua A. V. Blodgett⁵, John T. Whitteck^{1,2}, Gongyong Li¹, Satish K. Nair^{2,4}, Wilfred A. van der Donk^{1,2,3,4} & William W. Metcalf^{2,5}

Natural products containing phosphorus–carbon bonds have found widespread use in medicine and agriculture¹. One such compound, phosphinothricin tripeptide, contains the unusual amino acid phosphinothricin attached to two alanine residues. Synthetic phosphinothricin (glufosinate) is a component of two top-selling herbicides (Basta and Liberty), and is widely used with resistant transgenic crops including corn, cotton and canola. Recent genetic and biochemical studies showed that during phosphinothricin tripeptide biosynthesis 2-hydroxyethylphosphonate (HEP) is converted to hydroxymethylphosphonate (HMP)². Here we report the *in vitro* reconstitution of this unprecedented C(sp³)–C(sp³) bond cleavage reaction and X-ray crystal structures of the enzyme. The protein is a mononuclear non-haem iron(II)-dependent dioxygenase that converts HEP to HMP and formate. In contrast to most other members of this family, the oxidative consumption of HEP does not require additional cofactors or the input of exogenous electrons. The current study expands the scope of reactions catalysed by the 2-His–1-carboxylate mononuclear non-haem iron family of enzymes.

Hydroxyethylphosphonate dioxygenase (HEPD) was overexpressed in *Escherichia coli* with an amino-terminal hexa-histidine tag and purified by Ni²⁺-affinity chromatography. As isolated the protein was inactive, but in the presence of Fe(II) and O₂ conversion of HEP to HMP (Fig. 1) was observed by ³¹P NMR spectroscopy and high-performance liquid chromatography (HPLC). One molar equivalent of Fe(II) was sufficient for full reconstitution of activity. External reductants or cofactors were not required for catalysis, in contrast to most other non-haem Fe(II) oxygenases^{3,4}. Thus, all four electrons required for reduction of O₂ are provided by HEP.

The fate of the excised carbon was determined using synthetic HEP labelled with ¹³C at carbon 2 (2-¹³C-HEP). Stoichiometric amounts of HMP and a new product with a resonance of 171 parts per million (p.p.m.) in the ¹³C NMR spectrum were observed, corresponding to ¹³C-formate (Supplementary Fig. 1). Additionally, HMP formed with 1-¹³C-HEP and 2-¹³C-HEP was analysed by ³¹P NMR spectroscopy, resulting in a doublet (from ¹³C-HMP, Fig. 2a) and a singlet (from ¹²C-HMP) signal, respectively. The hydrogens at C1 are not removed in this process because the HEPD-catalysed oxidation of 1-²H₂-HEP produced 1-²H₂-HMP (Fig. 2a, inset).

The enzyme reaction was also performed in the presence of ¹⁸O₂ (99 atom per cent), with 2-¹³C-HEP as substrate to circumvent complications from spurious formate⁵. After derivatization of formate to its tert-butyldimethylsilyl ester, the products were monitored by gas chromatography–mass spectrometry (GC–MS), displaying ions at *m/z* 106 and 104 corresponding to loss of the tert-butyl group (M–57) from derivatized ¹⁸O,¹³C-formate and ¹⁶O,¹³C-formate, respectively. The

¹⁸O in formate exchanges with solvent in a time-dependent fashion in the protocol used (Supplementary Fig. 2), explaining the presence of the two products. The shortest exposure to the work-up and derivatization conditions resulted in >85% ¹⁸O-formate (Fig. 2b). Incorporation of ¹⁸O into HMP was assessed using liquid chromatography–MS (LC–MS). Approximately 60% of HMP contained ¹⁸O (*m/z* 115), with 40% containing ¹⁶O (*m/z* 113; Fig. 2c). This result was unexpected because the primary alcohol of HMP did not exchange under the reaction conditions. In an effort to understand the lower than expected ¹⁸O content in HMP, the reaction was also performed in the presence of 80% (v/v) H₂¹⁸O (95 atom per cent)/H₂¹⁶O. LC–MS analysis revealed 69% 2-¹⁶O-HMP and 31% 2-¹⁸O-HMP (Supplementary Fig. 3). When corrected for the ¹⁸O content of the labelled water, the two complementary experiments (¹⁸O₂ and H₂¹⁸O) are in good agreement and indicate the intermediacy of a species in which oxygen derived from O₂ exchanges with water. These results also demonstrate that HEPD is a dioxygenase. The results of all labelling studies are summarized in Fig. 2d.

HEPD does not have sequence homology with other proteins in the databases except for enzymes that probably catalyse the same transformation judging from their operon context. To gain three-dimensional information, the structure of HEPD was determined by

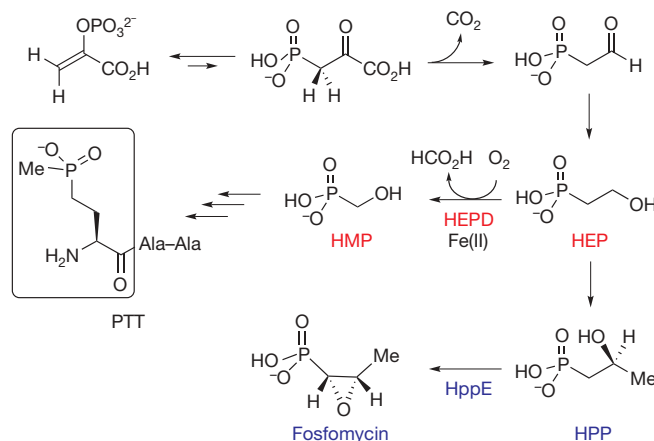


Figure 1 | Structure of phosphinothricin tripeptide and the reaction catalysed by HEPD. The biosyntheses of the commercial herbicide phosphinothricin (boxed in the phosphinothricin tripeptide (PTT) structure) and the clinically used antibiotic fosfomycin share several early steps starting with phosphoenolpyruvate before the pathways diverge after formation of HEP^{2,20}. HEPD catalyses the unprecedented conversion of HEP to HMP (for the steps from HMP to PTT, see Supplementary Fig. 8). The structurally related enzyme HppE converts HPP to fosfomycin²¹.

¹Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. ²Institute for Genomic Biology, ³Howard Hughes Medical Institute, ⁴Department of Biochemistry, and ⁵Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61802, USA.

*These authors contributed equally to this work.

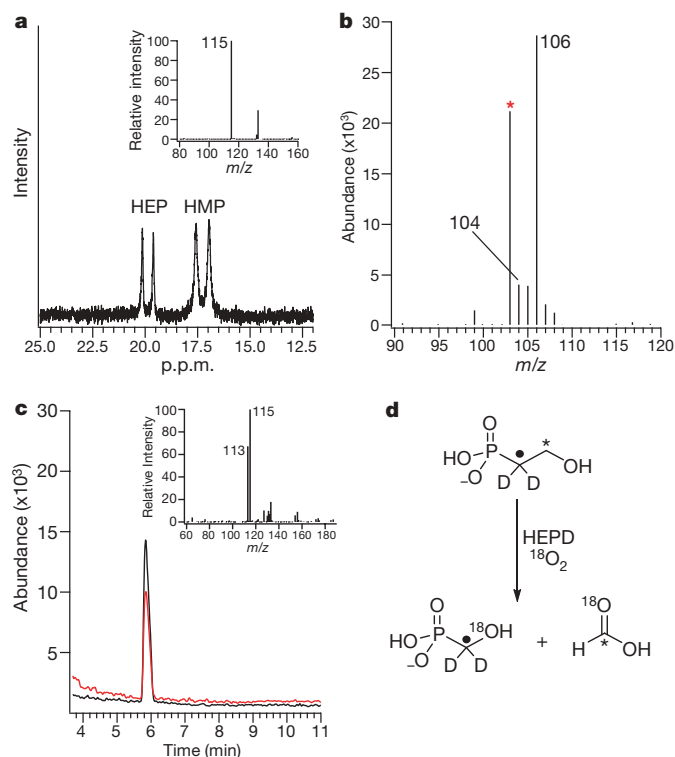


Figure 2 | NMR and mass spectral data from *in vitro* labelling studies. **a**, ^{31}P NMR spectrum showing production of ^{13}C -HMP (δ (chemical shift) 17.2 p.p.m.) from $1\text{-}^{13}\text{C}$ -HEP (δ 19.9 p.p.m.) after 50% conversion. Inset: mass spectrum of HMP derived from a reaction using $1,1\text{-}^2\text{H}_2$ -HEP. **b**, Mass spectrum of ^{18}O , ^{13}C -formate (m/z 106) produced by HEPD from a reaction using $^{18}\text{O}_2$ and $2\text{-}^{13}\text{C}$ -HEP. Spurious formate is denoted by an asterisk (m/z 103). **c**, Analysis of HMP produced from the HEPD reaction performed with $^{18}\text{O}_2$. Extracted ion chromatograms demonstrate that $\sim 60\%$ of the HMP contains ^{18}O (black line) whereas 40% contains ^{16}O (red line). Inset: summed mass spectrum from the total ion chromatographic peak. **d**, Summary of labelling studies from the HEPD reaction.

single-wavelength anomalous diffraction data collected from crystals of selenomethionine-incorporated protein, and refined to a Bragg limit of 1.8 Å. Crystallization was contingent on the presence of 50 mM CdCl_2 in the precipitant. The overall structure consists of imperfect tandem repeats of a bi-domain architecture (Fig. 3a). Each of the repeats is composed of an all- α -helical domain linked to a β -barrel fold characteristic of the cupin superfamily⁶. Despite the lack of appreciable sequence similarity, each of the two repeats is structurally homologous to the monomer of HppE⁷, a non-haem iron-dependent enzyme that converts (S)-2-hydroxypropylphosphonic acid (2-HPP) to fosfomycin (Fig. 1). However, the repeats of HEPD are not entirely discretely folded domains as the β -barrel domain of the first repeat is composed of interlacing β -strands from different parts of the molecule.

The disposition of the tandem domains of HEPD recapitulates many of the elements of the quaternary structure of HppE⁷, with the tandem arrangement found in HEPD structurally analogous to one-half of the HppE tetramer. This tetramer is stabilized through crossover interactions between the α -helical domains of the individual monomers. In contrast, the tandem arrangement in HEPD is held in place by two short orthogonal α -helices located at the junction between each helical and β -barrel domain (Fig. 3b). The β -barrel fold of the first repeat contains all of the canonical active site elements of non-haem iron enzymes, including a 2-His-1-Glu facial triad⁸. A Cd(II) ion is situated at the base of the active site and is coordinated by His 129, Glu 176 and His 182 (Fig. 3c). Well-defined spherical electron densities have been modelled as three water ligands forming a six-coordinate metal centre. The constellation of metal

ligands is similar to that of HppE, but the disposition of these ligands within the β -barrel is not conserved because Glu 176 of HEPD is situated on a different β -strand to the equivalent Glu 142 of HppE⁷. Furthermore, the spacing between the first two metal ligands in HEPD (HX_{46}E , where X represents any amino acid) is unique as these residues are closely spaced ($\text{HX}_{1-4}\text{E/D}$) in other facial triad enzymes⁷. In the second repeat in HEPD, the canonical 2-His-1-Asp/Glu is replaced by an arrangement of 2-His-1-Asn, and steric occlusion by residues Tyr 358 and Lys 404 precludes formation of a competent metal-binding pocket. Combined with the observed requirement for one equivalent of iron for full enzyme activity, the second repeat is probably vestigial and does not participate in catalysis. Given the structural similarity to the epoxidase HppE, HEPD was incubated with 2-HPP, the substrate for HppE. 2-HPP was converted to 2-oxopropylphosphonate (2-OPP), rendering the enzyme inactive in the process (Supplementary Fig. 5). No evidence was found for hydrogen peroxide formation in this transformation.

Attempts to produce crystals of Fe(II) -HEPD have failed owing to the high concentrations of Cd(II) required for crystallization. Crystals of SeMet-labelled Cd(II) -HEPD were grown in the presence of HEP and solved to a resolution of 1.92 Å, revealing electron density consistent with bidentate coordination of substrate to Cd^{2+} (Fig. 3d), which is similar to that observed for binding of 2-HPP to Co(II) -HppE⁷. In the HppE co-crystal structures, substrate binding induces substantial reorganization of the active site. In contrast, binding of HEP to Cd(II) -HEPD results only in the torsional movement of Tyr 98 about χ_1 (the torsion angle between the α - and β -carbons), resulting in a hydrogen bond interaction (2.3 Å) with one of the phosphonate oxygen atoms of the substrate. An additional hydrogen bond interaction occurs between N2 of Asn 126 (2.8 Å) and a different phosphonate oxygen.

Although oxidative scission of carbon-carbon bonds is well documented, previously characterized proteins typically act on substrates that contain aromatic³, alkene⁸ or 1,2-dihydroxy⁹ functionalities. No such activating groups are present in HEP, nor does HEPD need multiple successive oxidations such as in P450-mediated cleavage of the C20-C22 side chain of sterols¹⁰. Figure 4 presents a working model for the HEPD reaction. The substrate is proposed to bind to Fe(II) in a bidentate manner followed by reaction with O_2 , resulting in a $\text{Fe(III)}\text{-(O}_2^-)$ intermediate. This species is proposed to abstract a hydrogen atom from C2 of HEP to generate intermediate II, akin to similar steps proposed for isopenicillin N synthase and *myo*-inositol oxygenase^{9,11}. One important difference between isopenicillin N synthase and HEPD is that the former contains a sulphur ligand from the substrate that has been suggested on the basis of computational studies to stabilize the $\text{Fe(III)}\text{-(O}_2^-)$ intermediate and overcome the unfavourable energetics of O_2 binding and activation¹². Although a thiolate ligand is lacking in HEPD, the observation that external electrons are not required leaves no alternate pathways to convert intermediate I into a more reactive species found in other non-haem iron oxygenases⁴. Furthermore, HEP does not have low-energy electrons that can be provided by the substrate, as in the extradiol dioxygenases that use highly electron-rich catechol substrates. HEPD also does not contain other nearby metals that are important for activity, in contrast to the binuclear iron enzyme *myo*-inositol oxygenase¹³. Hence, the only option for HEPD seems to be hydrogen atom abstraction from C2 by intermediate I to produce the hydroperoxo complex II and a substrate radical.

Two different pathways can be envisioned to account for the labelling studies. The substrate radical could attack the hydroperoxide generating a ferryl species (Fe(IV)=O) and a hemiacetal (Fig. 4). The latter can undergo a retro-Claisen type C-C bond scission with the incipient negative charge on C1 attacking the electrophilic ferryl, potentially by means of a Horner-Wadsworth-Emmons-like intermediate stabilized by metal coordination. A ferryl would account for the required intermediacy of a species in which oxygen derived from O_2 can exchange with solvent¹⁴⁻¹⁹ (Supplementary Fig. 6). A second

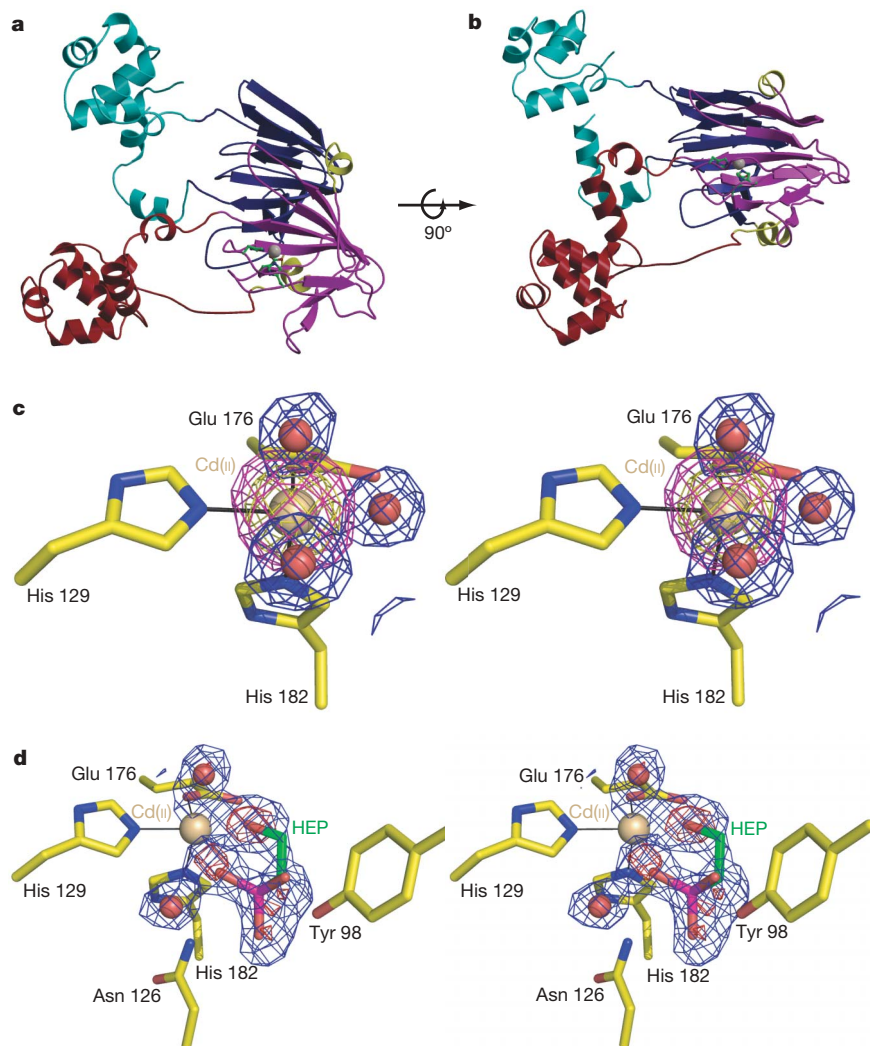


Figure 3 | Structures of Cd(II)-HEPD and Cd(II)-HEPD-HEP. **a, b**, Orthogonal views of Cd(II)-HEPD showing the tandem arrangement of the HppE fold including cupin domains (blue and purple), α -helical domains (red and cyan), cadmium (white sphere) and metal ligands (green). **c**, Stereoview of electron density maps using model phases ($F_o - F_c$). The first map (contoured at 3σ in blue) was calculated by omitting metal-bound solvents (red sphere) before one round of refinement. The second map, contoured at 5σ (purple mesh) and 14σ (yellow mesh), was calculated by omitting the cadmium (gold sphere) before one round of refinement. **d**, Stereoview of electron density maps ($F_o - F_c$) calculated using HEPD-HEP model phases. The map was calculated by omitting non-protein metal-bound ligands before one round of refinement (contoured at 3σ in blue mesh and 6σ in red). HEP carbons are shown in green.

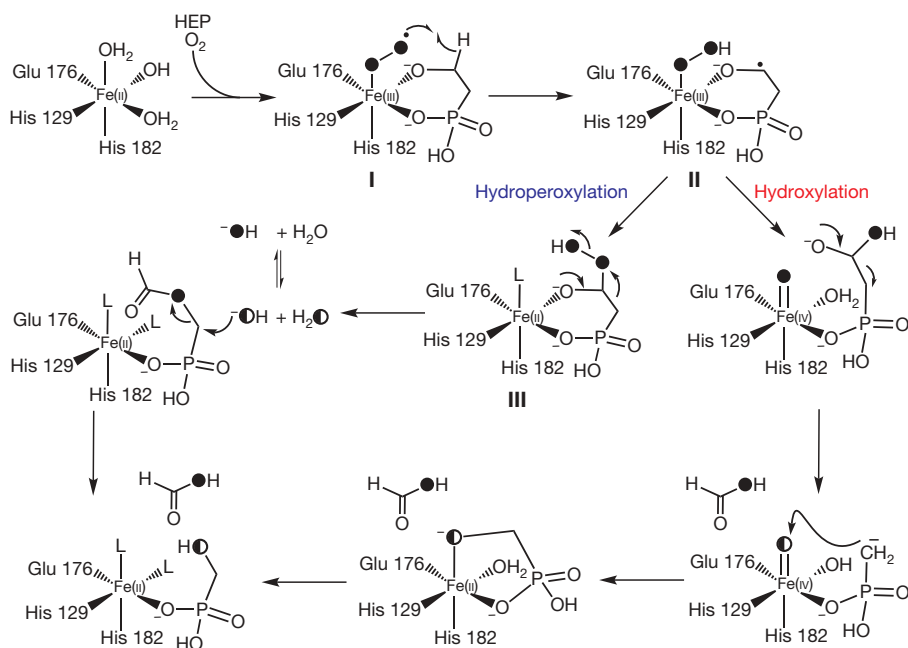


Figure 4 | Working models for the mechanism of catalysis by HEPD. The ferrous ion reacts with O_2 to generate the formal ferric-superoxide complex I. Hydrogen atom abstraction from C2 of HEP generates a substrate radical and ferric-hydroperoxide (II). From intermediate II two possible pathways can be envisaged. Radical attack on the ferric-hydroperoxide may cleave the O-O bond and produce a ferryl species and a hemiacetal (hydroxylation). C-C bond cleavage and attack of the resulting resonance-stabilized anion at the ferryl species generates formate and HMP. The weak hemiacetal ligand is posited to first dissociate from the iron to account for solvent exchange (Supplementary Fig. 6). Alternatively, intermediate II could be converted to the alkyl hydroperoxide III (hydroperoxylation), which can undergo a Criegee-type rearrangement to generate formyl-HMP. Hydrolysis must then occur at the methylene carbon to account for the oxygen labelling studies. Filled circles represent oxygen derived from O_2 and half-filled circles represent oxygens originally derived from O_2 that have exchanged with water. L, undefined metal ligand.

mechanism that can account for the observed data involves conversion of intermediate II to the hydroperoxide III, which can undergo a Criegee-type rearrangement to provide the formate ester of HMP (Fig. 4). Hydrolysis of this ester is expected to occur at the carbonyl carbon, but such a mechanism would not account for the incorporation of oxygen derived from solvent into HMP. This observation can be explained if hydrolysis took place via attack at C1 by solvent-exchangeable hydroxide released in the Criegee rearrangement. Both models can also explain the observed conversion of 2-HPP into 2-OPP, with the additional methyl group in the substrate inducing the cleavage of a C–O bond instead of a C–C bond to form a ketone (Supplementary Fig. 7). However, the lack of hydrogen peroxide production in this latter transformation favours the hydroxylation mechanism over the hydroperoxylation model. Further studies will be required to provide insights into the factors that result in epoxide formation by HppE and C–C bond cleavage by HEPD.

METHODS SUMMARY

HEPD was purified by immobilized metal affinity chromatography. Iron analysis²² on the as-isolated protein revealed iron content below the detection limits (0.05 equivalents) whereas fully active, Fe(II)-reconstituted protein contained 0.99 ± 0.05 equivalent of Fe per polypeptide. Typical aerobic assays contained, in a final volume of 1,000 μ l: 50 mM K⁺-HEPES, pH 7.5, 2–10 μ M HEPD, 5–10 mM HEP or isotopically labelled HEP. Formate produced was derivatized at 20 °C using 0.5 ml *N*-methyl-*N*-(tert-butyldimethylsilyl)trifluoroacetamide (MTBSTFA) plus 0.1% tert-butyldimethylsilyl chloride (TBDMSCl) for 20 min and monitored by GC–MS. HMP produced was analysed by LC–MS using atmospheric pressure chemical ionization (APCI). The stoichiometry of formate to HMP produced during the HEPD reaction was determined by ¹³C NMR spectroscopy. The potential for ¹⁸O incorporation from H₂O into HMP was evaluated by running the reaction in H₂¹⁸O. HEPD was anaerobically activated as described in the Methods. The mixture (150 μ l) was removed from the glove box and added to an aerobic mixture of H₂¹⁸O (832.6 μ l, 95 atom per cent) and HEP (17.4 μ l). The final concentrations of HEPD and HEP were 10.5 μ M and 3 mM, respectively. H₂¹⁸O was diluted to 79% in the 1 ml reaction. Intermittent bubbling of O₂ enabled the complete consumption of HEP over 1 h. HMP produced was analysed and a representative isotopic distribution of ¹⁸O/¹⁶O is shown in Supplementary Fig. 3. The reported percentage of ¹⁸O-HMP is corrected based on the amount of H₂¹⁸O in the assay.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 November 2008; accepted 4 March 2009.

- Seto, H. & Kuzuyama, T. Bioactive natural products with carbon–phosphorus bonds and their biosynthesis. *Nat. Prod. Rep.* **16**, 589–596 (1999).
- Blodgett, J. A. *et al.* Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide. *Nature Chem. Biol.* **3**, 480–485 (2007).
- Costas, M., Mehn, M. P., Jensen, M. P. & Que, L. Jr. Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. *Chem. Rev.* **104**, 939–986 (2004).
- Kovaleva, E. G. & Lipscomb, J. D. Versatility of biological non-heme Fe(II) centers in oxygen activation reactions. *Nature Chem. Biol.* **4**, 186–193 (2008).
- Shyadehi, A. Z. *et al.* The mechanism of the acyl-carbon bond cleavage reaction catalyzed by recombinant sterol 14 α -demethylase of *Candida albicans*. *J. Biol. Chem.* **271**, 12445–12450 (1996).
- Dunwell, J. M., Purvis, A. & Khuri, S. Cupins: the most functionally diverse protein superfamily? *Phytochemistry* **65**, 7–17 (2004).

- Higgins, L. J., Yan, F., Liu, P., Liu, H. W. & Drennan, C. L. Structural insight into antibiotic fosfomycin biosynthesis by a mononuclear iron enzyme. *Nature* **437**, 838–844 (2005).
- Grogan, G. Emergent mechanistic diversity of enzyme-catalysed beta-diketone cleavage. *Biochem. J.* **388**, 721–730 (2005).
- Xing, G. *et al.* Evidence for C–H cleavage by an iron-superoxide complex in the glycol cleavage reaction catalyzed by myo-inositol oxygenase. *Proc. Natl Acad. Sci. USA* **103**, 6130–6135 (2006).
- Lieberman, S. & Lin, Y. Y. Reflections on sterol sidechain cleavage process catalyzed by cytochrome P450(scc). *J. Steroid Biochem. Mol. Biol.* **78**, 1–14 (2001).
- Burzlauff, N. I. *et al.* The reaction cycle of isopenicillin N synthase observed by X-ray diffraction. *Nature* **401**, 721–724 (1999).
- Brown, C. D., Neidig, M. L., Neibergall, M. B., Lipscomb, J. D. & Solomon, E. I. VTVH-MCD and DFT studies of thiolate bonding to [FeNO]⁷/[FeO₂]⁸ complexes of isopenicillin N synthase: substrate determination of oxidase versus oxygenase activity in nonheme Fe enzymes. *J. Am. Chem. Soc.* **129**, 7427–7438 (2007).
- Xing, G. *et al.* Oxygen activation by a mixed-valent, diiron(II/III) cluster in the glycol cleavage reaction catalyzed by myo-inositol oxygenase. *Biochemistry* **45**, 5402–5412 (2006).
- Kikuchi, Y., Suzuki, Y. & Tamiya, N. The source of oxygen in the reaction catalysed by collagen lysyl hydroxylase. *Biochem. J.* **213**, 507–512 (1983).
- Baldwin, J. E., Adlington, R. M., Crouch, N. P. & Pereira, I. A. C. Incorporation of ¹⁸O-labelled water into oxygenated products produced by the enzyme deacetoxy/deacetylcephalosporin C synthase. *Tetrahedron* **49**, 7499–7518 (1993).
- Sabourin, P. J. & Bieber, L. L. The mechanism of alpha-ketoisocaproate oxygenase. Formation of beta-hydroxyisovalerate from alpha-ketoisocaproate. *J. Biol. Chem.* **257**, 7468–7471 (1982).
- Lindblad, B., Lindstedt, G. & Lindstedt, S. The mechanism of enzymic formation of homogentisate from p-hydroxyphenylpyruvate. *J. Am. Chem. Soc.* **92**, 7446–7449 (1970).
- Wackett, L. P., Kwart, L. D. & Gibson, D. T. Benzylic monooxygenation catalyzed by toluene dioxygenase from *Pseudomonas putida*. *Biochemistry* **27**, 1360–1367 (1988).
- Pestovsky, O. & Bakac, A. Aqueous ferryl(IV) ion: kinetics of oxygen atom transfer to substrates and oxo exchange with solvent water. *Inorg. Chem.* **45**, 814–820 (2006).
- Woodyer, R. D., Li, G., Zhao, H. & van der Donk, W. A. New insight into the biosynthesis of fosfomycin: discovery of the missing link illuminates the mechanism of methyl transfer. *Chem. Commun.* 359–361 (2007).
- Liu, P. *et al.* Protein purification and function assignment of the epoxidase catalyzing the formation of fosfomycin. *J. Am. Chem. Soc.* **123**, 4619–4620 (2001).
- Beinert, H. Micro methods for the quantitative determination of iron and copper in biological material. *Methods Enzymol.* **54**, 435–445 (1978).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Griffin, J. M. Bollinger, S. E. Denmark and T. Begley for discussions. This work was supported by grants from the National Institutes of Health (PO1 GM077596 to W.W.M., W.A.v.d.D. and S.K.N., and NIH RO1 GM59334 to W.W.M.) and by the University of Illinois.

Authors Contributions R.M.C. performed all biochemical assays shown, which were designed and analysed by R.M.C. and W.A.v.d.D. All structural studies were performed and interpreted by H.Z. and S.K.N. W.W.M. designed and J.A.V.B. performed initial biochemical reactions and identified the products. J.T.W. and G.L. synthesized all substrates. R.M.C., S.K.N. and W.A.v.d.D. wrote the manuscript.

Author Information Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession codes 3G7D (for the apo structure) and 3GBF (for the liganded structure). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.K.N. (s-nair@life.uiuc.edu), W.A.v.d.D. (vddonk@illinois.edu) or W.W.M. (metcalf@illinois.edu).

METHODS

Purification of recombinant HEPD. Recombinant HEPD was overexpressed as previously described². HEPD was purified by immobilized metal affinity chromatography using a Ni-NTA matrix (Qiagen). In a typical purification, 10 g of cells was thawed in 30 ml of buffer A (50 mM HEPES, pH 7.5, 300 mM NaCl, 20 mM imidazole, 10% glycerol) and allowed to incubate with lysozyme (1 mg ml⁻¹) for 30 min at 4 °C. Cells were lysed by two passages through a French press at a setting of 20,000 psi, and the lysate was centrifuged at 35,000g for 1 h at 4 °C. The supernatant was loaded onto a Ni-NTA column (1.5 × 3 cm) equilibrated in buffer A, and the column was washed with 10 column volumes of buffer A. The column was subsequently washed with buffer A containing first 50 mM and then 100 mM imidazole. The remaining bound protein was eluted with buffer A containing 250 mM imidazole. HEPD typically elutes between 100 mM and 250 mM imidazole and was pooled based on SDS-PAGE analysis. The protein was concentrated using an Amicon ultracentrifugation device with a 10 kDa membrane cutoff, and the buffer was exchanged to 50 mM HEPES, pH 7.5, 300 mM NaCl, 10% glycerol by gel filtration. Protein concentrations were determined by Bradford analysis using BSA as a standard. The protein was frozen in 100-μl aliquots and stored at -80 °C until ready for use.

HEPD activity assay. Anaerobic activation of HEPD was achieved in an anaerobic chamber obtained from Coy Laboratory Products, Inc. under an atmosphere of N₂ and H₂ (95%/5%). HEPD (30–70 μM) was treated with one molar equivalent of Fe(II)(NH₄)₂(SO₄)₂ for 20 min in a final volume of 200 μl. The solution was subsequently brought outside the glove box and combined with the remaining assay components. The reaction was initiated by addition of HEP (see Methods Summary for assay ingredients). For the quantification of formate, 150-μl aliquots of the assay mixture were removed at designated times and added to 2 μl of 9.2 M H₂SO₄ to quench the reaction. Precipitated protein was pelleted by centrifugation, and 50 μl of the supernatant was analysed by HPLC (Agilent Technologies 1200 series HPLC). Formate was chromatographed using an AMINEX HPX-87H (BioRad) column (300 × 7.8 mm) in an isocratic mobile phase (25 mM H₂SO₄). The flow rate and temperature were 0.6 ml min⁻¹ and 60 °C, respectively. Formate eluted at 14.1 min.

Formate analysis by GC-MS. Formate production was monitored on an Agilent 6890N (Agilent Technologies) gas chromatograph with helium carrier gas. Sample introduction was via split injection onto a HP-5 (5%-phenyl-methyl-polysiloxane) column (30 m, 0.32 mm inner diameter, 0.25 μm film thickness). The injection temperature was 250 °C. The initial column temperature was 40 °C, and was held for 5 min after injection before increasing to 230 °C at 15 °C min⁻¹. The temperature was held at 230 °C for the remainder of the 27 min program. The HEPD reaction was performed as described above. Sulphuric acid was added to a final concentration of 0.12 M to quench the reaction. The acidified solutions were subjected to three separate 0.5 ml diethyl ether extractions. The extracts were combined and formate was derivatized using

0.5 ml MTBSTFA plus 0.1% TBDMSCl (Pierce). Typical derivatizations were performed for 20 min at 25 °C. The solution (1–2 μl) was then directly injected into the GC-MS. Under the given GC conditions derivatized formate had a retention time of 7.4 min.

HMP analysis by LC-MS. Analysis and quantification of HMP was carried out by LC-MS using an Agilent 1200 series LC-MS equipped with a multimode electrospray ionization/APCI spray chamber. For HMP, APCI ionization was performed using positive ionization and chromatographic separations were achieved using a 150 × 4.6 mm Synergi C18 Fusion-RP column with a 4 μm particle size (Phenomenex). HMP has a retention time of approximately 6 min at a flow rate of 0.5 ml min⁻¹ using a 0.1% formic acid isocratic mobile phase. Injection volumes ranged from 5 μl to 15 μl. The nebulizer gas was N₂ at 8 l min⁻¹ with a nebulizer pressure of 40 psi. The drying gas temperature was 300 °C and the vaporization temperature was 250 °C. The capillary and charging voltages were both set to 2,000 V and the corona discharge current was 4 μA. The gain was set to 1.0 and the fragmentor voltage was 70 V.

Analysis of HEPD products by NMR spectroscopy. ¹H-coupled ³¹P-NMR and ¹H-decoupled ³¹P-NMR analyses were performed on a Varian Inova 600 spectrometer equipped with a 5-mm Varian 600DB AutoX probe tuned for phosphorus at 242.79 MHz. ¹³C spectra were recorded on a Varian Unity 500 spectrometer. Carbon chemical shifts were referenced to an external standard of 0.1% tetramethylsilane in CDCl₃. Phosphorus chemical shifts are reported relative to an external standard of 85% phosphoric acid (δ = 0). HEPD assays were performed as described above except that the reaction was terminated by removal of the protein by an ultrafiltration device (Millipore Amicon) equipped with a 10 kDa molecular weight cutoff membrane. D₂O was then added to the sample to a final concentration of 20% before data acquisition. The identification of the second product as formate was established by ¹³C NMR and GC-MS as seen in Supplementary Fig. 1a, b.

¹⁸O₂ incorporation assay. Assays using ¹⁸O₂ were carried out as described. All assay components were prepared and mixed anaerobically. 2-¹³C-HEP was substituted for HEP so that formate derived from substrate could be differentiated from spurious formate during GC-MS analysis. The glass vial was fitted with a tight rubber septum and brought outside of the glove box. The reaction was initiated by the introduction of 1 ml of ¹⁸O₂ (99 atom per cent, Isotec) via a gas-tight syringe. The solution was allowed to stir slowly at room temperature (25 °C) for 90 min at which time sulphuric acid was added to a final concentration of 92 mM to quench the reaction. After centrifugation to remove precipitated protein, HMP was directly analysed by LC-MS. Formate was extracted and derivatized as described above before GC-MS analysis. Time-dependent washout of the ¹⁸O label in formate, under acidic conditions, was observed by GC-MS (Supplementary Fig. 2) and has been noted elsewhere²³.

23. Shyadehi, A. Z. *et al.* The mechanism of the acyl-carbon bond cleavage reaction catalyzed by recombinant sterol 14 alpha-demethylase of *Candida albicans*. *J. Biol. Chem.* **271**, 12445–12450 (1996).

PROSPECTS

Post-PhD in an economic downturn

Despite the gloom, this isn't a bad time to begin a scientific career. Keep your nerve, Bart Noordam and Patricia Gosling advise.

Many graduate students finishing their studies this year will be facing their post-PhD career with some dismay. They may not have given much thought to the next step until recently because, just a year ago, opportunities seemed abundant.

Now, jobs are evaporating as the global financial crisis escalates. Reading about job losses and redundancies in the news may cause graduates to panic and jump at any offer, forgetting their career plans of a year ago.

But new PhDs should keep in mind that they are at the very beginning of a long career. They need to consider the pros and cons of the different options available and not panic.

It is important to understand the broader trends that influence graduates' current assessment of job prospects. The present economic situation has had little immediate impact on (semi-) governmental organizations such as national labs and universities. To keep the economy going, most governments will try to support and preserve jobs in the public sector as long as possible. So the number of vacancies for a postdoctoral fellowship — a typical first step after a PhD — probably will not shrink dramatically in the short term.

Moreover, the long-term job outlook for graduates with natural-sciences training is still excellent in many regions. Take the Netherlands, for example. An article last October in *Technomonitor* (<http://tinyurl.com/l63k6k>) reports a need during the next five years for an estimated 36,000 new academic employees trained in the natural sciences. The report is based on survey data from the Centre for Education and the Labour Market in Maastricht. Half of these new hires would replace existing employees (leaving for such reasons as retirement) and the other half is needed to meet demand based on expected growth. However, the data suggest that only 24,000 new graduates will be available to fill the vacancies.

A recent study commissioned by the German economics ministry found that young, skilled people are inadequately trained in academic subjects such as mathematics, natural sciences and technology. It estimates that Germany will lack up to 95,000 engineers and 135,000 scientists by 2014 (<http://tinyurl.com/ox6slt>).

In the long term, then, the Dutch numbers indicate that for every two graduates, three jobs are waiting. And estimated future vacancies could not be filled by the current workforce in either Holland or Germany. So



being a student in the natural sciences is good news. Indeed, some top economists believe the absence of a highly skilled workforce is a bigger long-term threat to the economy than the financial crisis.

No need to jump at the first safe job

In a financial downturn, newly minted PhDs might feel the understandable compulsion to overrate the importance of income and job security. They should worry about those later and make sure they do not accept a secure job that will limit career opportunities in the coming decades. Instead, they should seek a job in which they can acquire new work-related skills, both scientific and general,

and earn exposure to another field that may or may not be in science. Within a couple of years, the economy will probably blossom again, and employers will be desperately looking for qualified people. Then it will be time for scientists to cash in the work

experience and skill set they will have acquired in that less-than-ideal first job in order to achieve better prospects and a good salary.

Given the tight markets in some sectors, postdocs may face the temptation to stay in the same lab another year and continue their PhD research. Perhaps this means another chance at a scientific breakthrough unrealized during their PhD. Or perhaps there is a chance

to acquire additional skills. If so, this might be a convenient way to wait out the crisis. If, on the other hand, scientific advancement or acquisition of new skills seems unlikely, staying may be unwise and they ought to look elsewhere.

New graduates should keep in mind ways in which their first job might benefit their career, even if it is not their science 'dream job'. That first position in the working world will probably be significantly different from their PhD activities. But it is likely to involve working on complex problems that require a variety of skills. That first job probably will also have a major component that involves additional training in such 'soft skills' as giving presentations, writing concise e-mails, conducting meetings or making the most out of informal discussions at the coffee machine. Training on the job, in both hard and soft skills, is essential for advancing a career.

Economies are cyclic, with a crisis or recession coming along every seven to ten years. Most workers will probably face several crises during their careers. Young scientists seeking their first jobs are likely to be the most flexible, with few family constraints and a willingness and ability to move. So their first financial crisis should be easiest to cope with. ■

Bart Noordam is science faculty dean at the University of Amsterdam and Patricia Gosling is a senior manager at Amgen Europe in Zug, Switzerland. They are authors of *Mastering Your PhD* (Springer Verlag, 2006).

P. BARTON/CORBIS

Q&A

Blinded in 1984, geographer **Reginald Golledge** was this year named Faculty Research Lecturer by the University of California, Santa Barbara.



How did you begin on your career path?

In the 1960s, I discovered that a theoretical and quantitative revolution was transforming the previously descriptive field of geography, so I decided to research how people acquire spatial knowledge.

How did sudden blindness affect your career?

I was completely lost. I had no idea how I was going to teach without access to notes, prepared lectures or overheads. Figuring out how to continue my research was even more difficult. One day, two psychologists — Jack Loomis and Roberta Klatzky, both then at the University of California, Santa Barbara (UCSB) — asked if they could help. Roberta suggested I find ways to build on my previous mental-map research. They agreed to meet with me weekly and we began 25 years of intense collaboration, which took my research in a new direction towards dissecting spatial cognition. In the process,

I became more competent and was able to continue my academic life while helping other blind people around the world.

What is your greatest scientific achievement?

I pioneered a behavioural approach to human geography to study how humans find their way in the world. More than anything else, I think I opened the field's eyes to the fact that the geography you carry in your mind, your mental map and the way you process spatial information, are just as important as recording the facts of human existence on the surface of Earth. In addition to an objective reality, there is also subjective reality — what is stored in your mind's mental model of the world.

What has given you the most career satisfaction?

I've been developing a personal guidance system for blind travellers that allows them to be completely independent of

guides or guide dogs. Our prototype got a great deal of recognition, and now companies in many countries are producing these guidance systems. They're similar to vehicle-guidance systems, which use GPS and spatial databases or electronic maps.

Do you have any advice for disabled people who want to be scientists?

Disabled people can make long-term career plans, but it takes a strong commitment to their work.

What does this faculty award mean to you?

This is the best because it is given by your peers. It's the highest honour that UCSB faculty can award. It is so nice to know that what I've been doing has not gone unnoticed.

What is your motto?

You don't have to have sight to have vision. ■

Interview by Virginia Gewin.
Reginald Golledge sadly died on 29 May.

POSTDOC JOURNAL

Personal peer review



A goal all postdocs share is to publish their work so that they can build a competitive résumé. Recently I realized that the challenges of publishing a manuscript are strikingly parallel to many of the hurdles I will face when seeking a job. A few months ago I had an article published in *Genome Research* (B. J. Venters and B. F. Pugh *Genome Res.* **19**, 360–371; 2009). The process of getting this and other papers published taught me some important lessons about applying for jobs.

First, finding a journal suitable for my manuscript is analogous to finding a position for which I am qualified.

In other words, I seek to publish in a journal that is most relevant to my field and has the highest readership. However, if I misjudge the suitability of a manuscript for a particular journal, then the submission may be an exercise in futility. In the same way, whether in academia or industry, it is critical that I match my skills with the right position. Publishing papers requires tenacity, and this trait extends to mounting a successful job search.

Second, manuscript peer review is much like a job interview because my work is scrutinized and a decision is made one way or another.

During peer review, I want to persuade the reviewers that my work is novel and represents a significant contribution to the field. Similarly, in a job interview, my goal is to convince others that my unique blend of skills and experiences will be an asset to the university, institution or company to which I am applying. Two phrases I hope to hear in the future are, "Your manuscript has been accepted" and "You're hired". ■

Bryan Venters is a postdoctoral fellow at the Center for Eukaryotic Gene Regulation at Pennsylvania State University, University Park.

IN BRIEF

Stem cells in Maryland

A subsidiary of Korea-based biotech RNL Bio is expanding into Germantown, Maryland, where it is building a stem-cell research and development and manufacturing facility. The 930-square-metre building will be completed by the end of June, according to Donna Lee, director of business development, who says the company is hiring 50 lab technicians. By 2014, the company expects to double its floor space and hire stem-cell researchers. "There is a need for continuing studies on characteristics of the cells themselves and different disease indications," Lee says. RNL Biostar and parent company RNL Bio carry out research on therapies based on adult-derived stem cells.

Nuclear revival

The number of PhD degrees granted in nuclear engineering in the United States continues to increase, according to a survey by the Oak Ridge Institute for Science and Education in Oak Ridge, Tennessee, which identified a 72% jump — from 74 to 127 — between 2000 and 2008. The total rose 43% from 2007 to 2008, the survey found. Graduate enrolments have also increased each year since 2001, to 1,225 in 2008. "The promise of a nuclear renaissance has undoubtedly helped to drive increasing graduation and enrolment rates in nuclear engineering," says Eric Abelquist, a vice-president of Oak Ridge Associated Universities, which manages the institute for the US Department of Energy.

Minority health concern

The US National Center on Minority Health and Health Disparities is seeking to boost the diversity of the US biomedical workforce. The centre, part of the National Institutes of Health, has set up a two-year pilot research grant funded with \$500,000 of American Recovery and Reinvestment Act monies. Eligible research must be related to minority health or health inequalities and focus on diseases most often suffered by people in poor, rural or minority communities. The pilot programme could become permanent depending on its success, according to spokeswoman Kester Williams. It aims to funnel more biomedical researchers from diverse backgrounds into the field and increase the pool of those researching minority health or health disparities.

Timed release

Clean thoughts.

David Marusek

Neil is having a slow day at the office when Macie tweets him:

@NorthernGlory Booked Delta 289 tomorrow! Wheels down 5:03 AM. CANT WAIT TO SEE YOU!!! OXOXO

Neil twitters Macie an immediate reply:

@multicherry I CANT EITHER! See you at baggage claim. Crazy out of my head!

Neil spends the rest of the afternoon adrift in a sea of X-rated daydreams. Quitting-time never comes, and when it eventually does, he commutes home through a dense fog of resurging hormones. Somehow he arrives safely at their apartment door, only to be blindsided by reality when he opens it. Apparently, three months of poker nights, pay-per-view fights, World-of-Warcraft marathons, and everyday sloth have a way of adding up. Every room is an atrocity against good housekeeping — Trash, Filth, Stains and Odour. Neil scratches the back of his head and says: “Well, cat my dogs.”

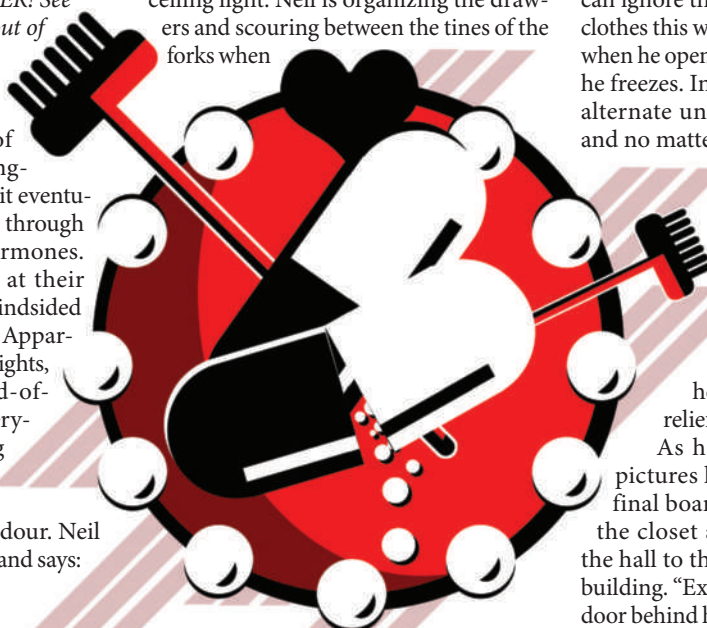
The performance aids section in the Walmart pharmacy is next to the colds/flu aisle.

Neil chooses a package of 12-hour, timed-release, maximum-strength Tidy Man gelpcaps. On the back label is a handy Tidy Man checklist of other Tidy Man products he might also need: furniture polish, oven cleaner, carpet steamer — it's a long list.

Back at home, Neil tosses his jacket into the hall closet and clears a spot on the couch. He downs a gelpcap with a swig of beer and lies back to do some calculating. It's 7:30 p.m., and he calculates that he has less than two hours per room before he has to leave for the airport. He realizes now that he should have picked up some dinner while he was out, and he gets up to see what he has in the fridge. Not much, it turns out — leftover lasagne well past its prime. He throws it away and empties the refrigerator, including the shelves and bins, and scrubs it inside and out, including on the top where dust and shopping bags seem to collect. He pulls the fridge away

from the wall and mops behind it and vacuums the condenser coils. It's only when he pushes it back into place and finds himself rearranging the door magnets that he realizes — Tidy Man is in the house.

Tidy Man is more frolic than work. Floors, oven, stovetop, dishes, windows, walls — it's all good. Scraping grease from the fan is good. So is shaking crumbs from the toaster and evicting dead flies from the ceiling light. Neil is organizing the drawers and scouring between the tines of the forks when



he realizes it's already past ten and he's still in the kitchen. This is housekeeping triage, and there are other rooms to save.

It's hard to leave, but Neil manages to charge into the living room with the laundry hamper and a carton of jumbo trash bags. He hums along with the vacuum cleaner. He strips the bed and flips the mattress. He wipes down the venetian blinds and houseplant leaves. He organizes DVDs and unsnarls the cords behind the entertainment centre. He polishes all of their shoes. The 40-gallon fish tank is a trove of fish poop, and the toilet and tub are more fun than a person should have alone.

As he's purging the medicine cabinet of its expired bottles and tubes, Neil makes the beginner's mistake of looking at himself in the mirror. What a shock. He sees nose hairs the size of fire hoses. He trims them back, which leads to eyebrow plucking, tooth bleaching, and fingernail and toe-

nail clipping. He brushes. He flosses. He shaves. He strips off his shirt and examines his back and shoulders in the mirror. He's reaching around with his electric razor when he is summoned by the clothes dryer — there's another load to iron and fold and put away.

It's 4:00 a.m. and time to leave, but it's hard to quit with so much left undone. Neil discovers that if he squints his eyes, he can ignore the worst of it. He changes his clothes this way and grabs his car keys, but when he opens the hall closet for his jacket, he freezes. Inside the closet it's like some alternate universe where chaos reigns, and no matter how hard he squints, Neil

is unable to pull away.

Precious minutes tick by, and with inspiration born of desperation, Neil speed-dials Macie and listens to the ringing, afraid it will go to voicemail, but no, she picks up. “What luck!”

he says with a little hiccup of relief. “Where are you?”

As he listens to her voice and pictures her standing wearily at her final boarding gate, Neil pulls out of the closet and squints his way down the hall to the front door and out of the building. “Excellent,” he says, shutting the door behind him. In the twilit distance, the sleepless city is rousing itself for another day. “I love you too. See you soon.”

Neil hangs up and bounds down the steps and dashes for his car with renewed energy. He has only yards to go when he spies a fast-food wrapper on the sidewalk. Don't pick that up, he tell himself, even as he bends over to pick it up. On the curb lies an empty beer bottle. He tries walking past it, but it's easier just to pick it up too.

The chain link fence has caught tattered sheets of newspaper. Indeed, all along the property line the fence is festooned with windblown litter.

Neil stands next to his car and consults his watch. He tries squinting the fence away, but in the rosy light of dawn, it's a gilded ribbon of garbage stretching like a promise as far as the eye can see. “Well, dog my cats,” he says and consults his watch again, recalculating the time of his release.

David Marusek keeps his cabin tidy in Fairbanks, Alaska, where he spins his quirky tales by the light of the midnight sun. Visit him at www.marusek.com.

JACEY